

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:34:30 ; Search time 6005.26 seconds
(without alignments)
8899.927 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtccaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pi.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Query Match	Length DB ID	Description

1	2524.2	98.8	2653	8	LES012798	AJ012798 Lycopersi
2	2524	98.8	2532	8	AF020390	AF020390 Lycopersi
3	1842.6	72.1	2629	8	AY029226	AY029226 Capsicum
4	1296	50.7	2653	8	AF064786	AF064786 Carica pa
5	1265.8	49.6	2643	8	AB046543	AB046543 Pyrus pyr
6	1256.2	49.2	2616	8	MAUBGALRP	L29451 Malus domes
7	1231.2	48.2	2783	8	CAAJ5042	AJ005042 Cicer ari
8	1218.8	47.7	2821	8	CAR011010	AJ011010 Cicer ari
9	1209	47.3	2166	8	AF229794	AF229794 Vigna rad
10	1207.4	47.3	2628	8	LAN011047	AJ011047 Lupinus a
11	1205.8	47.2	2628	6	A46218	A46218 Sequence 1
12	1205.8	47.2	2628	6	AR085663	AR085663 Sequence
13	1182.4	46.3	2704	8	FRX278705	AJ278705 Fragaria
14	1181.6	46.3	2202	8	AF004812	AF004812 Mangifera
15	1172.8	45.9	3010	8	LES012796	AJ012796 Lycopersi
16	1172.8	45.9	3038	8	AF154421	AF154421 Lycopersi
17	1171.2	45.9	2944	6	A46235	A46235 Sequence 18
18	1171.2	45.9	2944	6	AR085670	AR085670 Sequence
19	1168.2	45.7	2945	6	A44190	A44190 Sequence 2
20	1168.2	45.7	2945	6	AR029613	AR029613 Sequence
21	1168.2	45.7	2945	8	LEBETAGG	X83854 L.esculentu
22	1168.2	45.7	2989	8	LES012797	AJ012797 Lycopersi
23	1168.2	45.7	3205	8	AF023847	AF023847 Lycopersi
24	1146	44.9	3309	8	FRX278703	AJ278703 Fragaria
25	1132.2	44.3	3283	8	ATH270297	AJ270297 Arabidops
26	1109.2	43.4	3124	8	AY064690	AY064690 Arabidops
27	1108	43.4	2235	8	AY064690	AY064690 Arabidops
28	1108	43.4	2471	8	ATH270300	AJ270300 Arabidops
29	1108	43.4	2504	8	AY054589	AY054589 Arabidops
30	1105.6	43.3	3241	8	AB061017	AB061017 Persea am
31	1099.6	43.1	2221	8	ATH270308	AJ270308 Arabidops
32	1088.4	42.6	2476	8	AF367327	AF367327 Arabidops
33	1082	42.4	2442	8	ATH270298	AJ270298 Arabidops
34	1076.6	42.2	3152	8	ROBGNL	X77319 A.official
35	1004.8	39.3	2664	8	CAR012687	AJ012687 Cicer ari
36	986.4	38.6	2855	8	AY043231	AY043231 Vitis vin
37	958.4	37.5	2656	8	CAR6771	AJ006771 Cicer ari
38	952.4	37.3	2748	8	AY056285	AY056285 Arabidops
39	952.4	37.3	3067	8	ATH270299	AJ270299 Arabidops
40	936.8	36.7	2554	8	AF229795	AF229795 Vigna rad
41	934.2	36.6	3048	8	AF154424	AF154424 Lycopersi
42	924	36.2	2388	8	ATH270301	AJ270301 Arabidops
43	922.4	36.1	2451	8	AY069911	AY069911 Arabidops
44	920.8	36.1	2445	8	AY058098	AY058098 Arabidops
45	731.2	28.6	2781	8	ATH270304	AJ270304 Arabidops

ALIGNMENTS

RESULT	1
LES012798	LES012798 2653 bp mRNA linear PLN 07-JAN-1999
LOCUS	Lycopersicon esculentum mRNA for ss-galactosidase, clone tEG3.
DEFINITION	Lycopersicon esculentum mRNA for ss-galactosidase, clone tEG3.
ACCESSION	AJ012798
VERSION	AJ012798.1 GI:4138140
KEYWORDS	beta-galactosidase; ss-1,4-exogalactanase; ss-galactosidase; tEG3 gene.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum
REFERENCE	1 (bases 1 to 2653)
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
TITLE	Lycopersicon.
JOURNAL	Submitted (11-NOV-1998) de Silva J., Plant Sciences, Unilever Research, Unilever Research, Colworth Laboratory, Sharnbrook, Bedford, MK44 1LQ, UK
FEATURES	Location/Qualifiers
source	1. .2653
	/organism="Lycopersicon esculentum"

sig_peptide	/cultivar="Money maker"		Db	448		GAATGGAACCTTTGGGGGATTCCTCTGTTTGGCTAAAATATGTCCTGGTATGGAATTTAGA		507		
	/db_xref="taxon:4081"									
	/clone="tEG3"									
CDS	/dev_stage="Fruit"		Qy	481		acaaacaatcagccctttaaggttggtctatgcaaggattgttccagaaaaatagccaatg		540		
	91..159									
	/gene="tEG3"									
	91..2265		Db	508		ACAAACAATCAGCCCTTTAAGTGCGCTATGCGAGGATTTGTCAGAAAATAGTCAACATG		567		
	/gene="tEG3"		Qy	541		atgaagtgcagaaaaattgtttgaaatctcaaggaggaccaataataattgcccagatacaa		600		
	/standard_name="beta galactosidase"		Db	568		ATGAAGTCAGAAAAATTGTTTGAATCTCAAGGAGGACCAATAATAATTATGGCCCCAGATAGAA		627		
	/function="ss-1,4-exogalactanase"		Qy	601		aatgagatgcagcagtagaataaggaaaattggtgctcctggttaaagcttatatacaaaaatgg		660		
	/evidence=experimental		Db	628		AATGAGTATGGACCAGTAGAATGGGAAAATGGTGTCTCTGGTAAAGCTTATACAAAATGG		687		
	/product="ss-galactosidase"		Qy	661		gcagctcaaatggctgtgagtttgaaaaactggtgccatgcatgtgtccatgagatgtaagaacaag		720		
	/db_xref="GI:4138141"		Db	688		GCAGCTCAAAATGGCTGTAGGTTTGAAGAACTGGTGTCCCATGGATCATGTGTGAAGCAAGAG		747		
	/translacion="MLRNVLLLVICLLDFESSKVASVSDRAILLINGRKLILIS		Qy	721		gatgctcctgactcctgtgattgatacttgtaactgttaatggctcttactgcgaagggttccgctcct		780		
	SIHPRSTPQMPDLIOKAKDGLDVIETVFWNGHSPGKYNEGRYDLVRIKVV		Db	748		GATGCTCCTGATCCCTGTGATTGATTAATGCTTCTACTGCGAAGGGTTCCGTCTCT		807		
	QKAGLYVNRIGPIYVCAEWFGFPVWLKYYVPGMEFRNNQNPFFKVMHGFVOKIVNMH		Qy	781		aataagccttacaacacctaaaaatgtggacagaagtatggactggtgtgtatatacgaataatc		840		
	KSENLFEQSGPIYIAQIENIEYGPVEWIEGAPKAYTKWAAAGVLGKTVPWIMCKQ		Db	808		ATAAAGCCCTTACAAACCTTAAATGTGGACAGAAGTATGGACTGGCTGGTGTATACGAATTC		867		
	EDAPDVDTGNGFYCEGRPNKPYKPMWTEVMTGWTKEFGPIPORPAEDIAFSVA		Qy	841		ggtgtccaattccotcaagaccagcgaagacattgcatcttcagttcaggttcgaagttggt		900		
	RFVONGSFFNVYVHGTNFGRTSSGLFIATSYDYDAPLDEYGLLNEPKYVGHRLDLH		Db	868		GGTGTCCAAATTCCTCAAGACCACCCGAAAGACATTCGATTTTTCAGTTGCCAGGTTGT		927		
	KAIKLEPALVSSAAVTSLSGNOEAHVYRSKGACAAFLSNYDSRYSVKYTFQNPY		Qy	901		cagaaacaatggttcattcttcaattactacatgatatcatgaggaggaacaaattttggccgg		960		
	NLPWISILPDKTAVYNTACVNSQSSSIKMTFAGGGLSQSNEETPTADDSOTLT		Db	928		CAGAACAAATGGTTTCATTCTTCAATTACTACATGTATCATGGAGGAACAAATTTTGGCCGG		987		
	ANGWEQKNTVTRDSDYLWYMTNWNIAENEGFLKNGKDPYLTVMASAGHVLHFVNGKL		Qy	961		acatcatcagggtcttctcatgcaactagctacgattatgctgctcctctcgatgaatat		1020		
	SGTVYGLDNPGLKYSYGNVKLRAGINKISLSLSVSGVLPNVGVHYDTNWAGLVGPTLS		Db	988		ACATCATCAGGGCTTTTCATTGCAACTAGCTACGATTAATGATGCTCTCGATGAATAT		1047		
	GLNCSRLAKQKWSYKVLKSGESLSLSGSSVSWVWRSGLSVAQKPLTWYKATFN		Qy	1021		gggtgtcgtgaatgaaccacaaagtatgggcacttgagagacttacataaaagctatcaagcta		1080		
	APGNDPLALDMSMGKQIWIINGEGVRHWPVYIAGDCSKCSYAGTFNEKKKQCTNC		Db	1048		GGGTGCTGTAATGAACCAAAAGTATGGGCACCTTGAGAGACTTACATAAAGCTATCAAGCTA		1107		
	GOPSORWYHVPRLMPSGNNLVVFEWGGNPTGISLVRSR"		Qy	1081		tctgaacggcttttagtttcatcatatgctgoggtgactagtccttgggaagtaatacaagag		1140		
	/gene="tEG3"		Db	1108		TCTGAACCGGCTTTAGTTTCATCATATGCTCGGTGACTAGTCTTGGGAAGTAAATCAAGAG		1167		
	/EC_number="3.2.1.23"		Qy	1141		gctcatgtttatagatcaaaaatcctggagcttgctgctgttttttttccaaactatgactct		1200		
	/product="ss-galactosidase"		Db	1168		GCTCATGTTTATAGATCAAAAATCTGGAGCTTGCTGTGCTTTTATCCAACTATGACTCT		1227		
	/db_xref="taxon:4081"		Qy	1201		agatattcagtaaaagtcacaccttcagaaatagggccatacaactcctccatgggtccatc		1260		
	/function="ss-1,4-exogalactanase"		Db	1228		AGATATTTCAGTAAAAGTACCCTTTTCAGAAATAGGCCATACAAATCTGCCCTCCATGGTCCATC		1287		
	/evidence=experimental		Qy	1261		agcattctcccgcactgcacaaactgccgttttacaacactgcacaggtttaactctcaaacg		1320		
	/product="ss-galactosidase"		Db	1288		AGCATTTCTCCGACTGCAAAACCTGCCGTTTACAACTGCACAGCTTAACCTCAAAAGC		1347		
	/db_xref="GI:4138141"		Qy	1321		tcgagcataaagatgacgccttcaggttggtggtattccttggcagtcatacaaatgaagaa		1380		
	/translacion="MLRNVLLLVICLLDFESSKVASVSDRAILLINGRKLILIS		Db	1348		TCGAGCATAAAGATGACGCCTGCAGGTGTGGATTTGTCTTGGCAGCTACATAATGAAGAA		1407		
	SIHPRSTPQMPDLIOKAKDGLDVIETVFWNGHSPGKYNEGRYDLVRIKVV		Qy	1381		acgcctactgctgtagacagcgatcacacttacagctaacgactatgggaacagaaaaaac		1440		
	QKAGLYVNRIGPIYVCAEWFGFPVWLKYYVPGMEFRNNQNPFFKVMHGFVOKIVNMH		Db	1408		AGCCCTACTGCTGATGACAGCGATACACTTACAGCTTAACGGACTATGGGAACAGAAAAAC		1467		
	KSENLFEQSGPIYIAQIENIEYGPVEWIEGAPKAYTKWAAAGVLGKTVPWIMCKQ		Qy	1441		gtcacaagagatttcacagactatctgtggtacatgacaaaatgtaaaatagcatcctaat		1500		
	EDAPDVDTGNGFYCEGRPNKPYKPMWTEVMTGWTKEFGPIPORPAEDIAFSVA		Db	1468		gtcacaaagagatttcacagactatctgtgtggtgtacatgacaaaatgtaaaatgtatgactcct		1527		
	RFVONGSFFNVYVHGTNFGRTSSGLFIATSYDYDAPLDEYGLLNEPKYVGHRLDLH		Qy	1501		gaaggtattcttaagaacggaagatccttatctcactgttatgtcgcgttggtcatgtc		1560		

Db 1528 GAAGGATTTCTAAGAACGCGAAGGATCCTTATCTCTACTGTATGTCTCGCTGGTCATGTC 1587
Qy 1561 ttgatgttttcgtaacatggaacatcatcaggaactgtttatggtacattggataatcca 1620
Db 1588 TTGCATGTTTTCGTCATATGGAACACTATCAGAACTGTTTATGTTACATTGGATAATCCA 1647
Qy 1621 aaacttacatcagtggaactggaactgaagttaagagctggtatatacaagatttctctgctc 1680
Db 1648 AAACCTACATACAGTGGCAAGCTGAAGTTAAGAGCTGTTATTAACAAGATTCTCTGCTC 1707
Qy 1681 agtgttccgttgctctcccaactggtcgtgcattatgacatgacatggaatgcaggagt 1740
Db 1708 AGTGTTCCTGTTGCTCTCCCAAGCTTGGCGTGTATTAATGATACATGAATGCAGAGATT 1767
Qy 1741 ctagtccagtcacgttgagcggtctcaatgaaggtcaagaactgtgcgaacagaaaa 1800
Db 1768 CTAGTCCAGTCACGTTGAGCGGTCTCAATGAAGGTCACAAACTTGGCCGAACAGAAA 1827
Qy 1801 tggcttacaaggttgctctgaaaggcgaactcgtaagcttctactccttaagtggaggt 1860
Db 1828 TGGTCTTACAAAGTTGTTCTGAAAGCGGAATCGTTAAGTCTTCACTCCTTAAGTGGAGT 1887
Qy 1861 tctctgttgatgggttcaggttcactaatggtcctcaaaagcagccctgaacttggtac 1920
Db 1888 TCTTCTGTGATGGTTTCGAGGTTCTACTAGTGGCTCAAAAGCAGCCCTGACTGGTAC 1947
Qy 1921 aaggtcatttaacgcgcctggaggaaatgatcactagctttagacatggcaagtatg 1980
Db 1948 AAGGCTACATTTAAACGCGCTGGAGGAATGATCCACTAGCTTTAGACATGGCAAGTATG 2007
Qy 1981 ggaagagtcagatatggataaaatggtagagcgtagtgccttggcctggatatacata 2040
Db 2008 GGAAGAGTCAAGATGATGAATGATGAGGCGTAGTGCCTGAGGCTGGATACATA 2067
Qy 2041 gcacagagcagactcagcaaatcagattatctgagagcttcaacagagaaagtggga 2100
Db 2068 GCACAGCGAGCTCAGCAAAATGAGATTATGCTGGAAGCTTCAACGAGAAAGTGGCAG 2127
Qy 2101 actaacgcgcagcaactctcagagatggtaccatgttccacagatcgtggtcgaaccca 2160
Db 2128 ACTAATCGCGACAACTTCTCAGAGATGGTACCATGTTCCACAGTCTGGCTGAACCA 2187
Qy 2161 agtggaaactgttagtagtattcgaagaaatggggaggttaataccacaggaatttctcta 2220
Db 2188 AGTGGAACTTGTAGTAGTATTCGAAGATGGGAGGTAATCCACAGGAATTTCTCTA 2247
Qy 2221 gtcaggagatcaagataaagaactcgaagaatgaactgttccagtaactatggtgcttg 2280
Db 2248 GTCAGGAGATCAAGATAAAGAACTCGAAAAGTAAACTTGTTCAGTAACATGTTGCTTG 2307
Qy 2281 aattcgcgcgcaaaaaacacacacagactcaacaaatggaggtcacagtttgcgaattgc 2340
Db 2308 AATTGCGCGCAAAAATACATACACGAGCTTAACAATGGAGGCTACAGTTTGCAATGTC 2367
Qy 2341 agctgaataaacattagaagataaagaataatttattgataaaagagtatataaatttac 2400
Db 2368 AGCTGAATAAACAATTAAGATAAAGAAATATTGATTAAAGAGGATATATAAATTTC 2427
Qy 2401 agagaatttcttattcttctgtaaaacttggtttataaagttttatcacagaatttctg 2460
Db 2428 AGAGAATTTCTTATCTTTGTAACACTTGGTTTATAAAGTTTATACAGAATTTTCTG 2487
Qy 2461 ttatttgattatgagattgagaagattgtcacagcttcccaactattattagatacaaaa 2520
Db 2488 TTATTTGGATTATGAGATTGAAGAATGTTACAGCTTCCAAATACATTATTAGATACAAA 2547
Qy 2521 taaatttcaatgtaaaaaaataaaaaaataaaaaa 2553
Db 2548 TAAATTTTCATGTAACATATACATAATTATTACAA 2580

RESULT 2
AF020390

LOCUS AF020390 2532 bp mRNA linear PLN 25-JUL-2000
DEFINITION Lycopersicon esculentum beta-galactosidase (TBG4) mRNA, complete cds.
ACCESSION AF020390
VERSION AF020390.2 GI:7923928
KEYWORDS tomato.
SOURCE Lycopersicon esculentum
ORGANISM Lycopersicon esculentum
REFERENCE 1 (bases 1 to 2532)
AUTHORS Smith,D.L., Starrett,D.A. and Gross,K.C.
TITLE A gene coding for tomato fruit beta-galactosidase II is expressed during fruit ripening. Cloning, characterization, and expression pattern
JOURNAL Plant Physiol. 117 (2), 417-423 (1998)
MEDLINE 98289087
PUBMED 9625694
REFERENCE 2 (bases 1 to 2532)
AUTHORS Smith,D.L. and Gross,K.C.
TITLE A family of at least seven beta-galactosidase genes is expressed during tomato fruit development
JOURNAL Plant Physiol. 123 (3), 1173-1184 (2000)
MEDLINE 20349728
REFERENCE 3 (bases 1 to 2532)
AUTHORS Smith,D.L., Starrett,D. and Gross,K.C.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1997) HCQL, USDA/ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA
REFERENCE 4 (bases 1 to 2532)
AUTHORS Smith,D.L., Starrett,D. and Gross,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2000) HCQL, USDA/ARS, 10300 Baltimore Ave., Beltsville, MD 20705, USA
REMARK Sequence update by submitter
COMMENT On May 18, 2000 this sequence version replaced gi:3299895.
FEATURES
Location/Qualifiers
1..2532
/organism="Lycopersicon esculentum"
/cultivar="Rutgers"
/db_xref="taxon:4081"
/chromosome="12"
/map="R1 line: overlap of IL 12-2, IL 12-3"
/tissue_type="fruit"
1..2532
/gene="TBG4"
64..132
/gene="TBG4"
64..2238
/gene="TBG4"
/function="beta-galactosidase activity as measured by release of galactose from synthetic substrates such as p-nitrophenyl-beta-D-galactopyranoside; exo-galactanase activity as measured by release of galactose from a variety of native cell wall substrates"
/note="exo-galactanase; secreted protein; can release galactose from lactose"
/codon_start=1
/product="beta-galactosidase"
/protein_id="AAC25984.1"
/db_xref="GI:3299896"
/translation="MLRTNVLLLLVLCIDFFSSVKASVSDRAILLNKKRLILSG SIHYPRSTPQWPDLLQKADGGDLVITYVFNHGPSPGKYNFGRDVLRFKMY KENLFPESQGGPIIMAGIENEYGPVEIGAPGKAYTKAAQAQAVGLKTGPVIMCKQ EDAPDPVIDTCNGFYCEGFPNPKPKMTTEVMTGTYTKFGGPIQORPAEDIAFVA RFVONNGSPFNYYMHGTFNFTSSGLFATSYDYDAPLDEYGLLNEPKYGLRDLH KAIKLSEPALVSSAAVTSLSGNOEAHVFSKSCAACAEFLSNYDSYVKVTFQNPY NLFPWSISILPCKTAVYNTAQVNSQSSSMTTPAGGGLSNQSWSYNEETPTADDSDFLT ANGLEWQKNVTRDSSDLYWMTNNTASNEGFLKNGKDPYLTVMASAGHLVHVFVNGKL SGTVYGTLDNPKLTYSGNVKLKRAINKISLLSVSLVGLPNVGVHYDWTNAGVLGPVTL

gene

sig_peptide

CDS

GLNBSRLAKQKSYKVGLKESLSLSLSGSSSVWRGSLMAQKPLTWYKATFN
AGPNDPLALDMSMGKQIWIENGEGVRHWPYIAQGDCKSCSYAGTFNKKQTNIC
GQPSQWTHVPSRLKPSGNLLVVFEEWGNPTGISLVRSR"
BASE COUNT 800 a 431 c 559 g 742 t
ORIGIN

Query Match 98.8%; Score 2524; DB 8; Length 2532;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2527; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 aaaaaaagtccaatttttttctaaaaataaaaaaaatcoatttttttgaatgtaa 60
DB 1 AAAAAAGTTTCAATTTTTTCTAAAAATAAAAAAATTCATTTTTTTGAATGGAA 60
QY 61 aaatgctaagactaagtgtgtgtattagttattgtttttattgagattttttct 120
DB 61 AAAATGCTAAGACACATATGCTGTGTATTAGTTATTGTTATTGGAATTTTTTCT 120
QY 121 tcagtgaagctagtggtttcttatgatcacagagctataataatgggaagaata 180
DB 121 TCAGTGAAAGCTAGTCTTCTTATGATGACAGAGCTATAATCATAAATGGGAAAGAAA 180
QY 181 attctatttctggttcaattcaattatccaagaagcactccaagatgtggcctgatctt 240
DB 181 ATTCTTATTTCTGGTTCAATTCATTATCCAAAGACACTCCACAGATGGCCCTGATCTT 240
QY 241 atacaaaagcctaagatgagagcttagatgtattgaaactatgtttcttggaatgga 300
DB 241 ATACAAAAGGCTAAGATGAGAGGCTTAGATGTTATTGAAACTTATGTTTTCTGGAATGGA 300
QY 301 catgagccttccctggaaaaataattttgaaaggaagatatgatctttagattcatc 360
DB 301 CATGAGCCTTCCTCGGAAATATAATTTTGAAGGAAGATATGATCTTGTATGATTCATC 360
QY 361 aaaaaggtaacaagagcagggaacttatgtcaatttaactgtattggcccttaagctgtgct 420
DB 361 AAAATGGTCAAAAGAGAGGAGCTTATGTCAATTTACGTATTGGCCCTTAGCTGTGCT 420
QY 421 gaatggaactttggggatccctgtttggctaaaaatgtgcctgggtatggaatttga 480
DB 421 GAATGGAACCTTTGGGGATTCCTGTTGGCTAATAATATGTGCCCTGTGATGGAATTTAGA 480
QY 481 acaacaatcagccctttaagtggtgctatgcaaggatttggtcagaaataatgtaacatg 540
DB 481 ACAACAATCAGCCTTTAAAGTGGCTATGCAAGGATTTGTTCAAGAAATAGTCAACATG 540
QY 541 atgaagtcagaaaaatttggtaactccaaggaggagcaataattatggcccagatacaa 600
DB 541 ATGAAGTCAGAAAAATTTGTTGAAATCTCAAGGAGGACCAATAAATATGGCCAGATAGAA 600
QY 601 aatgagatggaccagtagaattgggaattgggtgctcctggttaagcttatacaaaatgg 660
DB 601 AATGAGTATGGACCAAGTAGAATGGGAATTTGGTGTCTCTGTTAAAGCTTATACAAAATGG 660
QY 661 cgagctcaaatgctgtagtttggaaactgggtgtcccatggatcatgtgtaagcaagag 720
DB 661 GCAGCTCAAAATGGCTGTAGTTTGAATACTGGTGTGCCATGATCATGTGTAAGCAAGAG 720
QY 721 gatgctcctgatcctgtgatgatactgttaagtgtctactgaggaaggttccgtcct 780
DB 721 GATGCTCCTGATTCCTGTATGATACTGTAAATGGCTTCTACTGCGAAGGGTTCGTCCT 780
QY 781 aataagccttacaacacctaaaatgtggacagaagtagtgactggctgggttatcacgaattc 840
DB 781 AATAAGCCTTACAAACCTAAAATGTGGACAGAGTATGGACTGGCTGTATACGAAATTC 840
QY 841 ggtgtccaattcctcaagaaccagccgaagacattgcattttcaagttggcaggtttgtt 900
DB 841 GGTGTCCAATTTCCPTCAAAGACCAAGCCGAACATTTGCAATTTTCCAGTTGCCAGGTTTGT 900
QY 901 cagaacaatggttcaattcctaactacatgtatcatgaggaacaaaatttggccgg 960
DB 901

DB 901 CAGAACAAATGGTTCATTTCTCAATTTACTACATGTATCATGGAGGAACAAATTTTGGCCGG 960
QY 961 acatcatcagggcttttcatgtcaactagctacgattatgatgtcctcctcctgatgaat 1020
DB 961 ACATCATCAGGGCTTTTCAATGCAACTAGCTACGATATGATGCTCTCTCGATCAATAT 1020
QY 1021 ggggtgctgaatgaaccaaagtatgggcacttgagagacttacataaaagctatcaagta 1080
DB 1021 GGGTGTCTGAATGAACCAAGATATGGCACCTTGAGAGACTTACATAAAGCTATCAAGCTA 1080
QY 1081 tetgaacgggttttagtttcatcatatgctgcggtgacttagcttgggaagtaatacaagag 1140
DB 1081 TCTGAACCGGCTTTAGTTTTCATCATATGCTGGGTGACTAGTCTTTGGAAGTAAATCAAGAG 1140
QY 1141 gctcatgtttatagatcaaaactctgagctgtgctgttttttttccaaactatgactct 1200
DB 1141 GCTCATGTTTATAGATCAAAATCTCGAGCTTGTGCTGCTTTTATCCAACTATGACTCT 1200
QY 1201 agatattcagtaaaagtccacttttcagaaataggccatacaactcctcccatggctccat 1260
DB 1201 AGATATTCTAGTAAAAAGTCACTTTTTCAGAAATAGGCCATACAAATCTGCCCTCCATGGTCCATC 1260
QY 1261 agcattctccgactgcaaaaactccgtttatacaacactgcaaggtttaactctcaagc 1320
DB 1261 AGCATTTCTCCGACTGCAAAACTGCCGTTTACAACTGCACAGGTTAACTTCTCAAAAG 1320
QY 1321 tcgagcataaagatgacgcctgcaggtgtgtgattctcttgccagctacatacaaatgaagaa 1380
DB 1321 TCGACATTAAGATGACGCCCTGCAGTGGTGGATTGCTTGGCAGTACATACAAATGAAGAA 1380
QY 1381 acgctactgctgtagcacagcgatacacttacagctaacgagactatgggaacagaaaaac 1440
DB 1381 ACGCTACTGCTGATGACAGCGATACACTTACAGCTAACGGACTATGGAAACAGAAAAAC 1440
QY 1441 gtcacaagagattcaatcagactatctgtgtgtacatgacaaaatgtaaatatagcatcaat 1500
DB 1441 GTCAAGAGAGATTTCATCAGACTTATCTGGTACATGACAAAATGTAATATATAGCATCTAAT 1500
QY 1501 gaagattctcaagaacggaaagatccttatctcactgttatgtcgcgtgggtcatgct 1560
DB 1501 GAAGGATTTCAAGAACGGAAAGATCCTTATCTCACTGTTATGTCGCTGGTCAATGTC 1560
QY 1561 ttgaatgtttcgtcaatggaaaaactatcaggaaactgtttatgtgtacatttggataatcca 1620
DB 1561 TTGCATGTTTTCGTCATATGGAAAACTATCAGAACTGTTTATGGTACATTTGGATTAATCCA 1620
QY 1621 aaactacatcagtgccaagtaagtttaagagctgtatttaacaagatttctcgtctc 1680
DB 1621 AAACCTACATCAGTGGCAACGTGAAGTTTAAAGCTGGTATTAAACAAGATTCTCTGCTC 1680
QY 1681 agtgttccgttggctcccgaaacgttggcgtgcattatgatataatgcaatgcaggagt 1740
DB 1681 AGTGTTCGTTGGTCTCCCGAACGTTGGCGTGCAATTTATGATACATGGAATGCAGAGTT 1740
QY 1741 ctaggctcagtcacgtttgagcgggtctcaatgaagggttcaagaaacttggcgaacagaaa 1800
DB 1741 CTAGGTCCAGCTCACGTTGAGCGGCTCTCAATGAAGGGTCAAGAAACTTGGCGAAACAGAAA 1800
QY 1801 tggcttacaaggttggctctaaaagcgaaatcgtttaagcttctcactccttaagtggagt 1860
DB 1801 TGGCTTTACAAGGTGGTCTGAAAGCGGAATCGTTTAAAGCTTCTACTCTCTTAAAGTGGAGT 1860
QY 1861 tctctgttgaatgggttcgaggttccactaaatggctcaaaagcagccctgacttgggtac 1920
DB 1861 TCTTCTGTTGAATGGGTTTCGAGGTTTCACTAATGGTCAAAAAGCAGCCCTGACTTGGTAC 1920
QY 1921 aaggtacatttaacgcgccttgaggaaaatgatccactagcttttagacatggcgaagtgt 1980
DB 1921 AAGGTACATTTAAGCGCCTTGGAGAAATGATCCACTAGCTTTTATAGATGGCAAGTATG 1980
QY 1981 ggaagagctcagatattgataaaatggtaagcgcttagtcgccatttgccttgggatacata 2040
DB 1981 GAAAAGGCTCAGATATGGATTAATGGTGAAGCGGTAGTGTGCCATTTGGCCCTGGATACATA 2040


```
Qy 2041 gcacaaaggcgactgcagcaaatgcagttatctgctgaaacttcaacagaaagaatgcgga 2100
|
|
|
Db 2041 GCACAAGCGACTGCAGCAATGCAGTTATCTGGAAGGTTCAACGAGAAGTGCAG 2100
|
|
|
Qy 2101 actaactgcggaacacctctcagagatggtaccatgcttccacagatggtgctgaaccca 2160
|
|
|
Db 2101 ACTAACTCGCGACACACCTTCTCAGAGATGGTACCATGTTCCACGATCGTGGCTGAAACCA 2160
|
|
|
Qy 2161 agtgaacactgttagtagtattcgaagaatggaggaatgaatccaacaggaatttctcta 2220
|
|
|
Db 2161 AGTGAACCTGTGTAGTAGTATTCGAGAATGGGAGGTAATCAACAGGAATTTCTCTA 2220
|
|
|
Qy 2221 gtcggagatcaagaataaagaaactcgaagaatgaaactgttccagtaactatggtgcttg 2280
|
|
|
Db 2221 GTCAGGAGATCAAGATAAAGAACTCGAAAGCTTAAACCTGTTTCAGTAATATGTTGCTTG 2280
|
|
|
Qy 2281 aatcgcgcgcaaaaatacatcacacgacgactcaacaaatggaaggtctacagtttgcaaatgc 2340
|
|
|
Db 2281 AATTCGCGCCGAAAAATACATACACGACGCTAACAAATGGAGGCTACAGTTTGCAAAATGCG 2340
|
|
|
Qy 2341 agctgaataaaacattagaagataaaagaataattgtattgataaaagagatataaatttac 2400
|
|
|
Db 2341 ACCTGAATAAACAATAGAAGATAAAGAAATATTGTATTAAGAGAGATATAAATTTTAC 2400
|
|
|
Qy 2401 agagaattttcttattctttgtaaaactttggtttataaaagttttacagaattttctg 2460
|
|
|
Db 2401 AGACAATTTCTTTATTCTTTGTAACACTTTGGTTTATAAGTTTATACACAATTTCTG 2460
|
|
|
Qy 2461 ttattggattatgagattggaagattgtcacagcttcccaactattatgaatacaaaa 2520
|
|
|
Db 2461 TTATTGGATTATGAGATTGAAGAAGATTGTACAGCTTCCAAATACTATTAGAATACAAA 2520
|
|
|
Qy 2521 taaatttcattgt 2532
|
|
|
Db 2521 TAAATTTCAATGT 2532
|
|
|
RESULT 3
LOCUS AY029226 2629 bp mRNA linear PLN 02-MAY-2001
DEFINITION Capsicum annum beta-galactosidase (BGI) mRNA, complete cds.
ACCESSION AY029226
VERSION AY029226.1 GI:13936235
KEYWORDS
SOURCE Capsicum annum.
ORGANISM Capsicum annum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
REFERENCE 1 (bases 1 to 2629)
AUTHORS Zainal-Abidin, W., Zainal, Z., Ali, Z.M. and Lazan, H.
TITLE A full length cDNA sequence of Capsicum annum L. beta
galactosidase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2629)
AUTHORS Zainal-Abidin, W., Zainal, Z., Ali, Z.M. and Lazan, H.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) School of Bioscience & Biotechnology,
Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia
FEATURES
source
1. .2629
/organism="Capsicum annum"
/cultivar="MC11"
/db_xref="taxon:4072"
/tissue_type="ripening fruit pericarp"
1. .12
/gene="BGI"
1. .2629
/gene="BGI"
13. .2187
/gene="BGI"
/codon_start=1
5'UTR
gene
CDS
```

Qy 772 ttccgctcctaataagccttacaacctaataatgtgacagaagtatgactggtggtat 831
Db 721 TTCCTCTTAATAGCCTTACAAACCGAAATGTGACAGAGTTTGGACGTGGT 780
Qy 832 acgaataattggtggttccattctcctcaagaccagcgcgaagacattgctttcagttgcc 891
Db 781 ACAAAATTTGGCGGTCCAAATTTCTCAACGGCCAGCTGAAGATATCGCGTTTTCAGTTGCA 840
Qy 892 aggtttgttcagaaacaatggttcattcttcattactactacatgatcatgaggaaacaat 951
Db 841 AGTTTGTTCAGAACCAACGGTTTCACTTCAATTTACTACATGATACACGGGGGAAACAAT 900
Qy 952 ttgtgcccgaacatcatcaggggttttttcattgcaactagctacgattatgatgctcctctc 1011
Db 901 TTTGCCGGGACATCTTCAGGGGTTTTTCATCGGACTAGCTATGATATGATGCTCCTATC 960
Qy 1012 gatgaatatgggttgcgaatgaacaaagatggtggcacttgagagacattacataaagct 1071
Db 961 GATGAGTATGGGCTTTCTGAATGAACCAAGTACGGGCACCTGAGAGAGTTGCATAAAGCG 1020
Qy 1072 atcaagctatctgaacggccttttagtttcaatcatatgctgcggtgactagttcttgaagt 1131
Db 1021 ATCAAGCAATGTGAACCGGCTTTAGTTTCATCGTATCCACGGTGACTAGTCTTGGTAGT 1080
Qy 1132 aatcaagaggctcatgtttatagatcaaaatctggagctgtgctgctttttttatccaac 1191
Db 1081 AATCAAGAGGCTCATGCTATAGATCGAAATCTGGAGCTTGTGCTGCAATCTCTGTCGAAC 1140
Qy 1192 tatgactctagatattcagtaaaagtcaccccttcagaaatagccaatacaatctgcctcca 1251
Db 1141 TATGATGCTAAATATTCAGTTCGAGCTCAGCTTTTCAGAAATTTGCCATACGATCTTCCTCCG 1200
Qy 1252 tggctcatcagcattctccgagctgaaaaactgcggttttcaaacacatgcacaggttaac 1311
Db 1201 TGGTCCATCAGCATTTCTCCGACTGCAAAACTGTGTATTAACACCCGCAAGGTTTAGC 1260
Qy 1312 tctcaagctcgagataaagatgacgctgcaggtggtggtattctctggcagtcatac 1371
Db 1261 TCCCAAGGCTCGAGCATAAAGATGACGCTGCAGGTGGTGGATTGTCTTGGCAGTCATAC 1320
Qy 1372 aatgaagaaacgctactgctgatgacagcgatatacaactacagctaaacgactatggaa 1431
Db 1321 AATGAAGACACACCGACTGCTGATGACAGCGATACACTTAGAGCTTAACGGGACTATGGAA 1380
Qy 1432 cagaaaaacgctcaagagattcatcacactatctgtgtgcatgacgaaatgtaaatata 1491
Db 1381 CAGAAACGTCACAGAGATTTTCATCGGACTATCTGTGTATACACAGCGTAAACATA 1440
Qy 1492 gcataataagaggatttctaaagacggaaggtactcttactactgtttatgtccgct 1551
Db 1441 GCATCTAACGAAGGATTTCTAAGAGCGGAAGGATCTTATCTCACGGTTATGTCGGCT 1500
Qy 1552 ggtcatgttctgcatgttttctgctcaatggaacactatcaggaactgtttatggtacattg 1611
Db 1501 GGTCAATGTTGTCATGTTTCTGCTAATGGAACCTTGCAGGAACGTATACGGGGCATTTG 1560
Qy 1612 gataatcaaaacttacaacagtggaacgtgaagttaagagctggtatttaacaagatt 1671
Db 1561 GATAATCAAAACTTACATACAGCGGCAATGTGAAGCTTAATGCGGGTATTAACAAGATT 1620
Qy 1672 tctctgctcaggttctcgttgggtctcccgcaactggtgctgcatatgatcatggaatt 1731
Db 1621 TCTCTCTCAGTGTTCGGTTCCTCCGAAACGTTGGCGGTACATATGATACGTGGAA 1680
Qy 1732 gcaggagttctaggtccagtcacgttgtagcggtctcaatgaagggtcaagaacttggcg 1791
Db 1681 GCAGGAGTTCTAGTCCAGTCACGTTGAGCGGCTCTCAATGAGGGGTACAGAGACTTGGCG 1740
Qy 1792 aaacagaatggtcttacaaggttgggtgaaagcggaatgtaagtcttcaactcctta 1851
Db 1741 AANCAAGAAATGGTCTATAAGAGTTGGTCTAAAGGCGGAATCGCTAAGTCTTTCACACGTTA 1800

Qy 1852 agtggagagttcttctgttgaaatgtcgaggttctactaaatggtcctcaaaagcagccctg 1911
Db 1801 AGTGTAGTCTCTCTCTGTCGAATGGGTTCAAGGTTTCACTGGTGGCTCGAACAACCCCTG 1860
Qy 1912 acttggtcacaggctcatatttaacgcgctggagaaatgatccactagcttttaacatg 1971
Db 1861 ACATGTTACAAGGCAACATTTAGCGCGCTGGAGGAAATGAACCCCTAGCTTTAGACATG 1920
Qy 1972 gcaagtatggaaaaaggtcagatatgataaattgtaagcgttagtgcgcatcctggcct 2031
Db 1921 GCGAGTATGGGAAAGGTCAGATATGGATCAATGTGAAGCGGTAGGTCGCCCTTTGGCT 1980
Qy 2032 ggatacatagcacaaaggcagctgcagcaaatgacgttatgctggaacgttcaacgagaag 2091
Db 1981 GGATACGACGACACAAGCGACTGTCAGCAAAATGACGTTCAGCTGGAACATTCAACGAGAAG 2040
Qy 2092 aagtgcgaaactaaactcgcggcaacacctctcagagatggtaccatgttccacgactcgtg 2151
Db 2041 AAATGTCAGACTTAACGTGGGACAGCTTCTCAGAGATGATCATGTCCCGCATCGTGG 2100
Qy 2152 ctgaaaccaagtggaaactgttagtagtatttcgaagaatgggaggttaataccaacagga 2211
Db 2101 CTGAAAACAAGTGGAAACTTGTGTAGTATTTGAAGAATGGGTGGTATCCCAACAGGA 2160
Qy 2212 atttctctagtcaggagatcaagaataaagaactcgaaagtaaaactgttccagtaacta 2271
Db 2161 ATTTCTTTAGTAAGAAGATCAAGATTAAGAGCTTCAAAAGT----- 2201
Qy 2272 tgggtgctgaattcgcgcgaaatacatcacagaactaaacataacgaatgaggtcagctacgttt 2331
Db 2202 -----AAAAATGCATACAGAAGCTAACAAATGAGGCTATA-TTT 2240
Qy 2332 gcaaatgcagctgaataaaacttagaagataaagaataatttgaataaaggagata 2391
Db 2241 GCAAAATGCAGCTGAATAATCAATTTGAAG-TACATAAATAATTGATAAAGGTATATAA 2299
Qy 2392 taaat 2396
Db 2300 AAAAT 2304

RESULT 4
AF064786 2753 bp mRNA linear PLN 13-NOV-1998
LOCUS Carica papaya beta-galactosidase precursor, mRNA, complete cds.
DEFINITION
ACCESSION AF064786
VERSION AF064786.1 GI:3869279
KEYWORDS
SOURCE papaya.
ORGANISM Carica papaya
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Caricaceae; Carica.
REFERENCE
1 (bases 1 to 2753)
AUTHORS Othman, R., Choo, T.S., Ali, Z.M., Zainal, Z. and Lazan, H.
TITLE A full-length beta-galactosidase cDNA sequence from ripening papaya
JOURNAL plant Physiol. 118 (3), 1102 (1998)
REFERENCE
2 (bases 1 to 2753)
AUTHORS Othman, R., Choo, T.S., Ali, Z.M. and Lazan, H.
TITLE Direct Submission
JOURNAL Submitted (12-MAY-1998) Biochemistry, Universiti Kebangsaan
Malaysia, Bangi, Selangor 43600, Malaysia
FEATURES
Location/Qualifiers
1..2753
/organism="Carica papaya"
/cultivar="Eksotika"
/db_xref="taxon:3649"
/clone="pBGL1"
/tissue_type="mesocarp"
/dev_stage="ripening fruit"
1..126
5'UTR
127..180 sig_peptide
127..2292 CDS

```
/EC_number="3.2.1.23"  
/function="involved in cell wall depolymerization"  
/codon_start=1  
/product="beta-galactosidase precursor"  
/protein_id="AAC77377.1"  
/db_xref="GI:3859280"  
/translation="MUKTNLIVLFLFCSWLSVEAVSYDHKAILIINGRRRIILSGSI  
HYPRSPQMPDLIQNAKEGGLDVIQTYFWNGHEPSPGNYFYEDRYDLVKFLKLVHQ  
AGLYVHLRISPIYICGEWNEGGFFWLKYYVPGIQFRTDNGPFPKQMQKFTKEIVMMKA  
EKLFEPQGGPIINSQENEXEYPIEWEIETGAPKAYTKWAAQMAVGLGTGVPWIMCKQED  
APDIDITONGFYENFMPNANYKPKMFTFAWTGWTEFGGPPVYPAEDPMWYSAVRE  
IQNRGSFIYMYHGTNFGRTAGPFIATSYDYDAPLDEYGLRREPCKWHLRDLHKT  
IKCESFLSVDPKVTSLGSNQEAHFVWTKISCAAFLLANDLYLSVRYTQNLPLDLP  
PWSVLILPDKTAVFYNTAKVYSQSLAKMIAVNSAFSWQSYNEETPSANDYVFTKDG  
LWQIISVTRDADLYMTDVTIGPDEAFILKNQDPIITVMSAGHALHVFVNGQLSGT  
VYGOLNPENKLFAGSKVLKRVGNKSVLSLAVGLPNVGLHFHFNAGLVLPVTLKGVN  
SGTWDMSKKWSYKIGLKGELSLHTVSSSSVWVEGSLLAORQPLIWKYTFNAPV  
GNPDIALDMSKMGQIWIWNGQISGRHWPGYKARGSCGCACNYAGIIDBKCHNSCGKA  
SQRWYHVPKSWLNPNTANLLVVFEEWGGDPTKISLVKRVV"  
mat_peptide  
3'UTR  
BASE COUNT 837 a 504 c 607 g 805 t  
ORIGIN  
/product="beta galactosidase"  
2293..2753  
181..2289
```

Query Match 50.7%; Score 1296; DB 8; Length 2753;
Best Local Similarity 74.9%; Pred. No. 2.1e-240;
Matches 1652; Conservative 0; Mismatches 545; Indels 9; Gaps 2;

Qy	24	taaaataaaaaaaattcatttttttgaattggaatggaataaaatgtaaggactaatgtgtt	83
Db	87	TGAAGGAGTAACAATAATTATTGATATTTTCATTTGCTATCATGTTGAAGACAAAACCTGGT	146
Qy	84	gttgtattagtattgtttattgatttttttttttttttttttttttttttttttttttttt	143
Db	147	CTTGTTCTTGTGTTGTTGTCATGG-----CTTTGTGCTGTTGAAGCTACTGTGTCCTTA	200
Qy	144	tgatgacagagctataatcaataaagggaagaaagaaattcttattcttggttcaattca	203
Db	201	CGACATAAAGCTATATCATTAATGCGCGAAGAGGATCTTATTTCTGGCTCCATTCA	260
Qy	204	ttatccaagaagactccacagatggtggcctgattcttatatacaaaaggctaaaagatgag	263
Db	261	TTATCCAGAGCACTCCTCAGATGTGGCCTGATCTTATACAAATGCTAAAGAGGAGG	320
Qy	264	cttagatttatgaaactattgttctggaatggacatgagccttctctggaataa	323
Db	321	GTTAGATGTCATACAGACTTATGTTTTTTTGAACGGACATGAGCCCCTCTCTCGAAATTA	380
Qy	324	taattttgaaggagatgatctgttagattcatcaaaatggatcaaaagagcaggact	383
Db	381	TTATTTTGAAGACAGGTATGATCTTGTAAAGTTTCATCAAGTTGTCATCAAGCTGGTCT	440
Qy	384	ttatgtcaattacgtattggcccttcacgtctgtgctgaaatggaaactttggggattccc	443
Db	441	GTAATGTCATCTCAGAATAAGTCCTTATATTGTTGGTGAATGGAATTTTGGGGGTTTTTC	500
Qy	444	tgtttgctaaaaatattgctgctgattggaatttagaacaacaataatcagccttttaagg	503
Db	501	TGTTTGCTCAATPACCTCTCTGTTATTCATTCAGACAGACATAATGGACCTTTCAAGGC	560
Qy	504	ggctatgaaggattttgttcagaaaaagtcacaatgatgaagtacagaaaaattgtttga	563
Db	561	ACAAATCAAAAATTTACAGAGAAATAGTCAACATGATGAAGCGACAGAAAGTTATTGA	620
Qy	564	atctcaaggaggaccataattatggcccgatatacaaaatgagtatggaccagttagaatg	623
Db	621	ACCTCAAGGGGGTCCAATAATTATGTCACAGATAGAGAAATGAGTATGGACCTATTGAGT	680
Qy	624	ggaattggtgctccttgtaagacttatcaaaaatggcagcctcaaatggctgtaggttt	683
Db	681	GGAAATTTGGAGCCCGGGGAAGCTTATACAAAATGGCGACAGAAAATGGCAGTGGGTCT	740

Qy	684	gaaactggtgtcccatgatcatgtgtaacaagagatgctcctgatcctgtgattga	743
Db	741	TGGCAGTGAGTCCCATGGATTATGCAACAAGAGAGATGCTCTTGACCCAAATATTGA	800
Qy	744	tacttgaatggcttctactcgcgaagggttcggtccctaataagccttacaacataaaat	803
Db	801	CACTTGCAATGGTTTCTATTGTGAAAAATTTTCATGCCAACGCCAACATCAACACCAAAAT	860
Qy	804	gtgacagaagatgaggactgctggtatatacaaaaatcgggtggtccaatctcctcaaaacc	863
Db	861	GTTTACAGAGCCCTGGACTGGCTGGTACACGGAATTTTGGCGGTCCAGTTCCTTTATAGACC	920
Qy	864	agccgaagacattgcatttttcagttgccaggtttgttcagaaacaaatggttcattcttcaa	923
Db	921	TGCAGAAGACATGCTTACTCCGTTGCAAGGTTCATTTCAGAAATAGGGGATCATTCATTAA	980
Qy	924	ttactaatgtatcataggagaaacaaaattttggcggagacatcaagggttttcaatgc	983
Db	981	TTATTATATGATACCATGGAGGAACAAAATTTTGGCAGAACTGCTGGAGGTCTCTTCATTGC	1040
Qy	984	aactagctacgattatgctcctcgcgatgaatatgggtgctgtaagtaaccaaagta	1043
Db	1041	TACTAGCTATGATTACGATGCCCTCTTGATGAGTATGGACTAAGGAGGGAGCCAAAATG	1100
Qy	1044	tgggcacttgagagacttacataaaagctatcaagctatctgaaacggcgttttagtttcac	1103
Db	1101	GGGCACTTGAAGGATTTGCATAAAACCATCAAAATTATGTGAACCATCTTTTAGTTCTGT	1160
Qy	1104	atatgctgggtgactagcttggaagtaatacaagggtcattgtttatagatacaaaatc	1163
Db	1161	TGATCTTAAAGTACATCGTTAGGAAGTAAACAAAGAGGCTCATGTGTTTGGACAAAAC	1220
Qy	1164	tggagcttgctcctcttttttccaaactatgactctagatatattcagtaaaagtcacott	1223
Db	1221	---CTCTTGTGCTGCATTCCTTGTGTAACCTAGATCTGAAGTACTCAGTAGAGTACACCTT	1277
Qy	1224	tcagaataggccatacaaatctgcctccatggtccatcagcattcttcccgactgcacaaac	1283
Db	1278	TCAAAACCTGCCTTATGACCTACCTCCTTGGTCTGTGAGCATTTCTTCTGAGTGCACAAAC	1337
Qy	1284	tggcgtttacaacactgcacagggttaactctcaaaagctcgagcataaaagatgcagcctgc	1343
Db	1338	TGTAGTTTCAACACTGCACAAAGTGTGTTTCAAGAGTTCGCTAGCAAAAGATGATGTGTGT	1397
Qy	1344	agggtggattctcttgccagtcatacaatgaagaaacgcctactcgtgtagcacgga	1403
Db	1398	CAACAGTGCATTCCTTGGCAGTCTGACACGAAGAAACACCTTCCGCAAAATTTATGATGC	1457
Qy	1404	tacacttacagctaacggaactatgggaacagaaaaacgtcacagagattcatcagacta	1463
Db	1458	TGTATTTACCAAAAGATGGCTGTGGGAACAGATAAAGTGTACCAGAGATGCTACAGATTA	1517
Qy	1464	tctgtgtacatgacaaatgtaataatagatctaaagaaagatttctaagaagacggaaa	1523
Db	1518	CTTGTGTATATACAGATGTGACATAGGTCCCTCATGAAGCACTTCTTGAGAAATGGGCA	1577
Qy	1524	ggatccttatcctactgtatgtccgctggtcgtctgctgctgctgttttcgtaagtaaa	1583
Db	1578	AGATCCCAATTTTGACAGTCATGTCAGAGGCCATGCTTTTGTCATGTTTTTGTGAATGTCA	1637
Qy	1584	actatcaggaaactgttatgggtacattggataatccaaaacttacatacagtggaacgt	1643
Db	1638	ACTATCAGGAACCTGATATGACCAATTTGGAAAAATCCCAAACTAGCTTTTAGTGGCAAGT	1697
Qy	1644	gaagttaagagctggtaataacaagatttctcgtcagtglttccggttggtctccogaa	1703
Db	1698	GAAACTGAGAGCAGAGTCAACAAAGGTTTCTTTACTAAGTATCGCTGTGGCTTCCGAA	1757
Qy	1704	cgttggcgctgattatgatacatggaatgcaaggattcttaggtccagtcacgttgagcgg	1763
Db	1758	TGTTGGCTTACACTTTGAAACATGGAATGCTGGGGTTCCTGGGTCCAGTACATTTGAAAGG	1817

```
Qy 1764 tctcaatgaagggtcaagaacttggcgacaacagaaatgggttcttcaagggttggtctgaa 1823
Db 1818 GGTGAATTCAGCAACATGGGATATGTAATGGAATGGTCTTACAAAGATGGTCTGAA 1877
Qy 1824 agggaaatcgtaagcttcaactcctaagtggaagttctctgttgtaagtggttcgagg 1883
Db 1878 AGGCCAAGCCCTTAGCCCTTCATACAGTTAGTGGCAGTTCGTGTGTAGTGGGTGAAGG 1937
Qy 1884 ttcaactaatggctcaaaaagcagccctgacttggtcacaaggtcatttaacgagcctgg 1943
Db 1938 ATCATTACTAGCTCAAGACAAACCCTCATTTGGTACAAAGCTACTTTTACGCACCACT 1997
Qy 1944 aggaaatgatccactagcttttagacatggcgaagtatgggaaaggtcagatatgataaa 2003
Db 1998 AGGTAAATGATCCATTAGCTTTAGATATCAACAGTATGGAAGGATGGAATGATGATAA 2057
Qy 2004 tggtagaagcgtaggtccatggcctggatatacatagacacaaagcgactcagcaaatg 2063
Db 2058 TGGTCAAAAGTATTGGACGCCCACTGGCCCTGGATATAAAGCTCGTGAAGTTGTGGCTTG 2117
Qy 2064 cagttatgtggaactgtcaacgagaagaagtgcggaactgaactcgcgacacactctca 2123
Db 2118 CAACCTACTCGTGAATATATGATGAGAAAATGTCATAGTAACGTGGAAAGGCTCTCA 2177
Qy 2124 gagatggtaccatgttccacgatcgtggtcgaacaaagtggaacttggtagtagtatt 2183
Db 2178 GAGATGTTACCATGTTCTCGCTCGTGCCTCAACCAACTCGAACCATTAGTTGTTTT 2237
Qy 2184 cgaagaatggaggaagtaaccacaggaattctctagtcaggaga 2229
Db 2238 TGAAGATGGGGTGTGTATCCCAACAAAGATTCTTTTGGTGAAAGA 2283

RESULT 5
AB046543 AB046543 2643 bp mRNA linear PLN 27-JAN-2001
LOCUS Pyrus pyrifolia pJPBGAL mRNA for beta-D-galactosidase, complete
DEFINITION
AB046543
AB046543.1 GI:12583686
ACCESSION
VERSION
KEYWORDS
SOURCE
Pyrus pyrifolia (cultivar:Housul) ripening fruit cDNA to mRNA,
clone:pJPBGAL.
ORGANISM
Pyrus pyrifolia
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Rosales; Rosaceae; Maloideae; Pyrus.
REFERENCE
1 (sites)
Tateishi,A., Inoue,H., Shiba,H. and Yamaki,S.
Molecular cloning of beta-galactosidase from Japanese pear (Pyrus
pyrifolia) and its gene expression with fruit ripening
Unpublished
2 (bases 1 to 2643)
Tateishi,A., Inoue,H., Shiba,H. and Yamaki,S.
Direct Submission
Submitted (25-JUL-2000) Akira Tateishi, College of Bioresource
Sciences Nihon University, Lab. of Pomology and Vegetable Crops
Science, 1866 Kameino, Fujisawa-city, Kanagawa 252-8510, Japan
(E-mail:tateishiebrs.nihon-u.ac.jp, Tel:81466843622,
Fax:81466843622)
FEATURES
Location/Qualifiers
source
1..2643
/organism="Pyrus pyrifolia"
/cultivar="Housul"
/db_xref="taxon:3767"
/clone="pJPBGAL"
/tissue_type="fruit"
/dev_stage="ripening"
183..2378
/gene="pJPBGAL"
183..2378
/gene="pJPBGAL"
/codon_start=1
```

```
/product="beta-D-galactosidase"
/protein_id="BAB21492.1"
/db_xref="GI:12583687"
/translation="MGVGIOTMWSILLFSCIPSAASVSYDHKAILINGOKRILLIS
QSIHYPRSTPEMPPDLJOKADGGLDIQYVFWNGHEPSPGKYFYFEDRYDLVFKIL
VQAGLEVNLRIGIPYCAEWNFGFVWLAKYVPGIAFTDNEPEKAAQKRFKIVSM
KBAKLFQOTGGPIILSOIENEFGFVEWEGAGKAYTKWAAQMAVGLDTGVPWIMCK
QEDAPDVIDTCNGFYCENFKPKNDYKPMWTEVWTGYTEFGGAVTRPADVAFVS
ARFTOSGSEALVNYVMYHGGTNGFRTAGGPMATSYDYDAPLDEYGLLRPKWGLHRLD
HKAFTKCESALVSDPSVTKLGSNOEAFVKSDESDFLAFLANDYAKYSVKVSGGGQY
DLPPWSISILPDKTEVYSTAKVGSQSSQVQMPFVHSGFPWQSFIEFTTSSDEDTT
LDGLYEINIRDTDYLWYMTDITIGSDAEFLKNGKSPLLTIFSAHALNVFINGQL
SGTVYSGLENPKLFSQNVNLRSGINKLALLISVGLPNVGTTFETWAGHNPITLK
GLNSGTWDMGSMGWTKYTKGLKEALGLHTVTGSSVSEWVEGSPSKMAKQPTLWYKATFN
APPCDAPLALDMGSMGKGTWINGSVGRHWPGYIARGSCDCSYACTYDDKKCRTHC
GEPSORWYHIPRSWLTPGNLLVVFEEWGGDPSRISILVERGTALDAKKP"
polyA_signal 2599..2604 622 g 704 t
BASE COUNT 785 a 532 c 622 g 704 t
ORIGIN
Query Match 49.6% Score 1265.8; DB 8; Length 2643;
Best Local Similarity 74.0%; Pred. No. 1.4e-234;
Matches 1619; Conservative 0; Mismatches 567; Indels 3; Gaps 1;
Qy 113 tttttcttcagtgaaagctagtttcttctatgatgacagactataatcaataatgga 172
Db 235 TTTTCTTCGACGTTTCGGCTTCTGTGAGTTACGACCACAAAGCTATAATAATTAATGGC 294
Qy 173 aaagaaaaattcttattcttggtcaattcattatccaagaagcactccacagatggc 232
Db 295 AGAAAGGATTTTAAATTTCTGGCTCCATCTACTATCCCAAGGACACTCTCAGATGTGGC 354
Qy 233 ctgatctatacaaaagcctaaagatgaggtcttagattgataaacttatgtttct 292
Db 355 CGATTTTAATTCAGAGGCCAAAGATGGAGCTTGGATGTTATACAGACCTATGTGTTT 414
Qy 293 ggaatggacatgagcctctctcctggaaaaataataatttgaaggagaatgatcttgta 352
Db 415 GGAATGGCCACGAACCTTCTCCGGGAAATAATATTATTCGAGGACAGATATGATTGGTCA 474
Qy 353 gattcataaagaatggtacaaagacagagactttatgcaatttcagttatggccctacg 412
Db 475 AGTTTCATCAAGCTGGTGCACAACAGCGCCTATTGTTTAATCTCCGGATTGGCCCTTATG 534
Qy 413 tctgtcgaatgaactttgggattccctgttttgctaaataatgtgcttgatgg 472
Db 535 TTTCGCTGAAATGAACATTCGGGGATTCCCGAGTTTGCTGAAATATGTCCTCGAATCG 594
Qy 473 aatttagaacaacaataacagccttttaaggtggctatgcaaggatttggtcagaaaaatag 532
Db 595 CTTTTCGAACGCAATGAGCCCTTCAAGGCGCAATGCAAAAATTTACACAGAAGATTG 654
Qy 533 tcaacatgatgaatcagaaaaattgttgaaatcctcaaggaggagcccaataattatggccc 592
Db 655 TCACATGATGAAGGCGAGAGAAGCTGTTCAAACTCAAGAGAGGCTCTATAATTTCTCTCTC 714
Qy 593 agatacaaaaatgagtatggaccagtagaattggaaattggtcctcctgggtaaaagtata 652
Db 715 AGATAGAAAATGAATTTGGACCTGTGGAATGGGAAATTTGGTCCGCTGGAAAAAGCTTACA 774
Qy 653 caaatgggcagctcaaatggctgtaggtttgaaaaactgggtgccccatggatcatgtgta 712
Db 775 CCAATGGCAGCTCAGATGGCTGTAGGCTAGACACTGGAGTTCATGGATTATGTGCA 834
Qy 713 agcaaggagatgctcctgatcctgattgattgatacttgaatgcttactcgcaagggt 772
Db 835 AGCAAGAGGATGCCCCCGATCCCTTTATGACACTTGAATGGTGGTTTCTACTGTGAGAAAT 894
Qy 773 tccgtcctaataaagccttacaaacctaaaaatgtggacagaagatggaactggtggtgata 832
Db 895 TCAAGCCAAATAGGACTATTAAGCCCAAAATGTGGACAGAACTCTGGACTGTGGTATATA 954
```


SGTVYVGSLENPKLFSFQNVNLRSGINKLALLISVGLPNVGTHTFETWAGVLGPITLK
GLNSGTWDMSSWKYTYTGLKGEALGHTVSSSVEWEGPSSMAEQPLTWYKATFN
APPDAPLALDMGSMGQIWIINGOSVGRHWPVGIARGSCDGSYATDDBKCRTHC
GEPQSRWHIPRSWLTPTGNLLVVFEEWGGDPSSILVERGTALDAKKL"
mat_peptide 244..2316

/evidence=experimental

/product="unnamed"

2371..2616

/note="putative"

2616

BASE COUNT 763 a 521 c 623 g 709 t

ORIGIN

Query Match 49.2%; Score 1256.2; DB 8; Length 2616;

Best Local Similarity 74.9%; Pred. No. 1e-232;

Matches 1586; Conservative 0; Mismatches 528; Indels 3; Gaps 1;

Qy	113	ttttttctcagtgaaagctagtggtttcttattgatgacagagctataatcataatggga	172
Db	227	TTTTTCTGCGAGCTTCAGCTTCGTGAGTTATGACCACAAAGCTATAATAATTAAATGGGC	286
Qy	173	aaagaaaaattcttattctggttccaattcattatccaaagaagcaactccacagatggc	232
Db	287	AGAAAAGGATTTTAATTTCTGGCTCCATTCACATATCCAGAAAGCACTCCCGAGATGTGGC	346
Qy	233	ctgatcttatcaaaaggctaaagatgagcgcttagatgttattgaaacttatgtttct	292
Db	347	CGGATTTAATTCAGAAAGCCAAAGATGAGGCTTGGATGTTATACAGACCTATGTGTTTT	406
Qy	293	ggaatggacatgagccttctctctggaaataataattttgaaggaagatatgatcttgta	352
Db	407	GGAATGGCCATGAACCTTCTCGGGAAATATTATTTCGAGGAAAGATATGATTTGGTCA	466
Qy	353	gattcatcaaaatggtacaaagagcagagcaactttatgcaaatcaattgagcccttacg	412
Db	467	AGTTTATCAAGCTGTGTGCAACAAGAGCCCTATTGTTTAATCTCGGATTTGGCCCTTAG	526
Qy	413	ctgtgctgaatggaaactttggggatccctgtttgtgctaaatatgtgctggtatgg	472
Db	527	TTTGGCTGAATGGAACTTCGGGGGATTTCCCGATTTGGCTGNAATATTGTCCTGGAATCG	586
Qy	473	aatttagaacaacaatcagcctttaaagggtggtcattcaaggattgttcagaaaaatag	532
Db	587	CTTTTCGAAGCGACAATGAGCTTCAAGGGCGCAATGCAAAAATTTACAGAGAAGATTG	646
Qy	533	caacatgatgaagtcagaaaaattgtttgaatctcaaggaggaccaataattatggccc	592
Db	647	TCAGCATGATGAAGCGCAGAGAAGCTGTTCAAACTCAAGGAGGTCCTATAATTCTCTCTC	706
Qy	593	agatacaaaatgagtagtaccagtagaatgggaatttggtcctctggttaaaagcttata	652
Db	707	AGATAGAAAATGAATTTGGACCTGTGGAATGGGAATTTGGTGACCTGGAAAAGCTTACA	766
Qy	653	caaaatgggcagctcaaatggcgttaggtttgaaactgggtgtcccatggatcatgtgta	712
Db	767	CCAAATGGCGAGCTCAGATGGCTGTAGGCTAGACACTGGAGTTCCATGGATTATGTGA	826
Qy	713	acgaagagatgctcctgatactggtgattgatacttgaatggtcttactcaggaaggt	772
Db	827	AGCAAGAGGATGCCCCGATCCGTTATGACACTTTGCAATTTGCAATTTCTACTGTGAAAT	886
Qy	773	tcctgcctcaataagccttcacaaacctaaatgtggacagaagtatggactggctgtgata	832
Db	887	TCAAGCCCAATAAAGGACTATAGCCCCAATAATGTGGACAGAGCTCTGGACTGGTGTGATA	946
Qy	833	cgaaatctgggtgtccaattcctcaaaagcagccgaagacattgatcttccagttgcca	892
Db	947	CAGAATTCGGTGGGCGAGTTCCCACTAGACCTGCAAGAGATGTGGCATTTTCAGTTGCTA	1006
Qy	893	ggtttgttcagaaacaatggttcattcttcaattcaatcatcatgatgaggaacaatt	952
Db	1007	GGTTTCATAAAGCGGCGGTTCGTTTTTGAACATATTACATGTATCCACCGGAGGACGAAT	1066

Qy	953	ttggcggacatcatcaggggttttccattgcaactagatgattatgatgctcctctcg	1012
Db	1067	TTGGCCGAACAGCGGAGGTCCCTTCATGCGCCACTAGCTATGACTACGACGCCCTTAG	1126
Qy	1013	atgaatatgggttgctgtaataaacaagtatgggcacttgagagactataaagcta	1072
Db	1127	ACGAATATGGACTACCCGGGAACCAAAAGTGGGACATTTGAGAGATCTGCACAAAGCA	1186
Qy	1073	tcaagctatcgaacggcgttttagttccatcatatgctgcggtgactagtgtgaagta	1132
Db	1187	TTAAATCATGTGAGTCTGCTTTAGTGTCCGTGTGACTTCCTAGTGACTAAACTCGGAAGTA	1246
Qy	1133	atcaagagctcatgttttatagatcaaaatctggagctgtgctgtcttttttatccaat	1192
Db	1247	ATCAAGAGGCTCATGTATTCAANTCAGAGTC--GGATTGGCTGCATTCCTCGCAAAAT	1303
Qy	1193	atgactctagatattcagtaaaagtcacctttcagaataggccatacaaatctgctccat	1252
Db	1304	ATGACGCAAAATACTCTCTTTAAAGTGAGCTTTGGAGCGGGCAGTATGACCTGCCGCAT	1363
Qy	1253	ggtccatcagcattcttcccaactgcaaaactgcccgtttacaacactgcacaggttaact	1312
Db	1364	GtCCATCAGCATTCCTCCGACTGCAAAACCGAAGTTTACAACACTGCAAHGTTGGTT	1423
Qy	1313	ctcaaaagctcagacataaagatgacgctgcaggtgggtggattctctctggcagtcataca	1372
Db	1424	CGCAAGCTCGCAAGTTCAGATGACACACAGTACATAGTGGATTCTTCTGGCAGTCATTCA	1483
Qy	1373	atgaagaaacgctactctgatgacagcgatacaacttaacagctaacagactatgggaac	1432
Db	1484	TCGAAGAAACACACGCTTCTGTATGAGACCGGATACAACTACATTTGACGGGTGTATGAGC	1543
Qy	1433	agaaaaacgtcacaaagatcactcagactatctgtgtacatacaaaatgtaaatatag	1492
Db	1544	AAATAAATATCACTAGGGATATCAGACGACTTCTGTGTACATGACAGATATCACAAATAG	1603
Qy	1493	catctcaatgaaggtattctaaagaacggaaagatcccttatctcactgttatgtccgctg	1552
Db	1604	GTTCTGATGAAGCGTTCTAAAGAAGCGAAAGTCCCCACTTCTTACGATCTTTTCAGCAG	1663
Qy	1553	gtcagtcttgcactgtttctgcaatggaaactatcaggaactgtttatggtacattgg	1612
Db	1664	GTCATGCTTGAATGTTTTCATCAATGTGTAGTACATGAGAACCTGTTATGGGTATCTGG	1723
Qy	1613	ataatccaaaacttacatacagtgcaagtgcaagtgaaagttaagagctggtattacaagattt	1672
Db	1724	AGATTCCTTAATTTATCATTCAGTCAAAACGTTGAACCTGAGATCTGGCATCAACAACCTTG	1783
Qy	1673	ctctgctcagtggttccggttggtctcccgaaacggttggtggtgcatttatgatacatggaatg	1732
Db	1784	CATTGCTTTAGCATTTCCGTTGCTGTCCGGAATGTTGGTACTCACATTTGAGACATGGAACG	1843
Qy	1733	caggagtctcaggtccagtcacggtgagcgtctcaatgaagggttcaagaacttggcga	1792
Db	1844	CGGGAGTTCTAGGCCCGATCACGTTGAAGGCTCTGAATTCAGGAACATGGGACATGTGAG	1903
Qy	1793	aacgaaatgctcttacaaggttggtctgaaagcgcaatcgtaagtcttcaactctaa	1852
Db	1904	GGTGAATAGACGTACAAGACTGGTCTGAAGAGTGGAAGCTTTAGGCCCTCCACTACTGTTA	1963
Qy	1853	gtggaggtcttctgttgaatgggttccaggttcaactaaggttcaaaagcagcccttga	1912
Db	1964	CTGGAGGTTCTTCTGTAATGGGTAGAAGGAGCCATCGATGCTGAAACAAACCCCTTA	2023
Qy	1913	cttgtaacaaaggtacatttaacgcgctgaggaataatgataccactagcttttagacatgg	1972
Db	2024	CATGTTACAAAGGCTACTTTTAATGCACCACAGGCTGATGCTCTCCATTTAGATTATGG	2083
Qy	1973	caagtatgggaaaaaggtcagatagataaatggtgaagcgtaggtgcgcatgtgctg	2032
Db	2084	GAAGCATGGGAAAAGGTCAGATATGGATTAATGGACAGAGCGTGGGACGCCACTGGCCCG	2143

Qy	891	caggtttgttcagaaacaaatggttcattcttcaattactaacatgatatcatgatgagaacaaa	950
Db	992	AAGATTATACAAAATTCGCGGTCTATGTTAACTACTATATGTACCATTGTGGAACATAA	1051
Qy	951	ttttgccggacatacatcagcggttttcattgcactagctacgatttatgatgctctct	1010
Db	1052	TTTTGGAGGCATCTGCTGGCCTTTTCATTGCCACAAGCTATGACTACGATGCACCTAT	1111
Qy	1011	cgatgaatatgggttgtctgaatgaaccaaagttagtggcacttgagagacttacataaagc	1070
Db	1112	TGATGAATACGGACTTCTTAAGTGAACCAAAATGGGGCATCTAAGAAAAATTTACATAAAGC	1171
Qy	1071	tacaagctatctgaaccgggtttagttttcoatactatgctgcggtgactagctctggaag	1130
Db	1172	AATAAGCAATGTAGCCGATTTTAGTGCTGTGGATCCGACGGTGTCTAGGCCAGGAAA	1231
Qy	1131	taatcaagaggctcatgtttatagatcaaaaaactgagagctgtgctgttttttatccaa	1190
Db	1232	AAACCTTGAGGTACATGTTTACAACAGCAGCACCTGGTGCCCTGCCGTGCATTCCTTCGAAA	1291
Qy	1191	ctagactctagatattcagtaaagtcacacctttcagaataggccatacaaatctgcctcc	1250
Db	1292	TTATGACACCACATCTCCGGCAAAGTTACATTTGGAAATGGCCAATAGGACCTTACCGCC	1351
Qy	1251	atggtccatcagcattctcccagctgcaaaactgccgttttacaacactgcacaggt---	1307
Db	1352	TTGGTCTATCAGTATTTCCTGACTGCAAAACTGCTGTTTTCAACACTGCAAAAGGTTGG	1411
Qy	1308	tasctctcaaaagctcgagcataaagatgacgcctgcaggtggtggattctcttggcagtc	1367
Db	1412	TACAGTACCAAGCTTCCATAGAAAAGATGACTCCGTGTGAGTAGCGCATTCGATGGCAGTC	1471
Qy	1368	atacaatgaagaaagcctactgctgatgacagcgatacacattacagctaacggaactatg	1427
Db	1472	ATACAAGCAAGCACCCGCTCGTCCGGTATAGATGATTCCACCACAGCAAAATGCACPTTT	1531
Qy	1428	ggaaacagaaaacgtccaagagattcatcagactatctgtggttacatgacaaatgtaaa	1487
Db	1532	GGAGCAGATCAAGGTGACTCGCGATTCCTCAGATTATTGTTGGTATATGACAGATGFCAA	1591
Qy	1488	tatagcatctaataagaggatttctaaagaaacggaagagatccttctcactgttatgctc	1547
Db	1592	CATTAGTCTCTAATGAAGGTTTTATAGAATGGTCAGTATCCTGTTCTTACTGCAATGTC	1651
Qy	1548	cgtcgtcatgtcttgcattgttttcgttcaaitygaaaactatcaggaactgtttatggtagc	1607
Db	1652	AGCTGGCCATGTTCTGCATGTTTTTTGTTAAATGGTCAATTTTCGGGAACTGCTTATGGAGG	1711
Qy	1608	attggataatccaaactacatacagfggcaacgtagaagtaaagactcggtattaaacaa	1667
Db	1712	ATTGGGAATCCTAAATTAACATTTAGTAATAGTGTAAAGCTGAGGTTTGGAAATTAACAA	1771
Qy	1668	gatttctctgcagtggtttccgttgggtctccgaacggttggcgtgcatttatgatacatg	1727
Db	1772	GATTTCTTTACTAGTGTGGCGTCCGGTCTCTCGAATGTTGGCTTGCACATATGAACATG	1831
Qy	1728	gaatgcaggagtcttagtccagtcacgcttgagcggctctcaatgaaggggtcagaagaactt	1787
Db	1832	GAATGTGTGTGTTAGGCCCGGTAAACGTTGAAGGGTCTTAAACGAGGGGACAAAGACATT	1891
Qy	1788	ggcgaacacagaatggtcttaacaggttggctcaaaaggcgaaatcgttaagttctcactc	1847
Db	1892	GTCGTGACAGAATGGTCTTACAGATTGGACTGAAAGGGGAAACCTTGAACCTTTCATAC	1951
Qy	1848	cttaagtggaagtctctctgttgaatgggttcgaggttccaataatggctcaaaagcagcc	1907
Db	1952	CTTAATFGGAAGTCTCTGTTCATATGACAAAAGGATCATCGTGTGTGTGAAAAAACAACC	2011
Qy	1908	ctgtacttggtacaaggctacatttaacgcgcttggaggaaatgatccactagctttaga	1967
Db	2012	TTTGACTTGGTACAAGGCAACTTTGACGCGCCCGCTGGCAAGCATCCATTTGGCGGTAGA	2071

Qy	1968	catggcaagtatgggaaaaagctcagatatagataaaatggtgaagcgtagtgcgccattg	2027
Db	2072	TATGAGTAGCATGGGAAAAGGTGAAATATGGTGAATGGTGAAGACATTTGTCGCGCATTTG	2131
Qy	2028	gcctggatcacatgacaaagcgactcagcaaaatgcagttatctggaacgttcaacga	2087
Db	2132	CCCGGCATATATACGACGTGGAAGTTGTGGTGTGTGTAATTTATCCGGGAACGTTTACCGA	2191
Qy	2088	gaagaatgcggaaactaaactcggacaacctctctcagagatggtaccatgtctccacgac	2147
Db	2192	TAAAAAATGCCGAAACAGCTGTGGACACACCCACCAAAAAATGGTATCATATTTCTCGCTC	2251
Qy	2148	gtgcttgaacccaatggaaacttgtagtagtatttcgaagaatggggaggtaataccaac	2207
Db	2252	CTGGGTGAACCAAGAGGGAACCTCTTAGTGTGTGCTTGAAGAATGGGAGCGCATCTTCT	2311
Qy	2208	aggaattctctagtcaggagatcaaga	2235
Db	2312	TGGAATTTCATTGGTGGAAGAACAATGA	2339
RESULT	8		
CAR011010			
LOCUS	CAR011010	2821 bp	mRNA linear
DEFINITION	Cicer arietinum mRNA for beta-galactosidase, clone CanBgAl-4.		
ACCESSION	AJ011010		
VERSION	AJ011010.1	GI:3641864	
KEYWORDS	beta-galactosidase.		
SOURCE	chickpea.		
ORGANISM	Cicer arietinum		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Cicereae; Cicer.		
REFERENCE	1 (bases 1 to 2821)		
AUTHORS	Labrador,E.		
TITLE	Direct Submission		
JOURNAL	Submitted (14-SEP-1998) Labrador E., Dpto. Fisiologia Vegetal, Univ. Salamanca, Campus Miguel de Unamuno. Pza. Doctores de la Reina s/n, 37007-Salamanca, SPAIN		
REFERENCE	2 (bases 1 to 2821)		
AUTHORS	Dopico,B., Esteban,R. and Labrador,E.		
TITLE	The fourth beta-galactosidase clone found in Cicer epicotyls		
JOURNAL	Unpublished		
FEATURES	Location/Qualifiers		
source	1..2821		
	/organism="Cicer arietinum"		
	/cultivar="Castellana"		
	/db_xref="taxon:3827"		
	/clone="CanBgAl-4"		
	/tissue_type="etiolated epicotyls"		
	/clone_lib="CAN-5"		
	/dev_stage="5 days old seedling"		
	204..2375		
	/EC_number="3.2.1.23"		
	/codon_start=1		
	/product="beta-galactosidase"		
	/protein_id="CAA09457.1"		
	/db_xref="GI:3641865"		
	/db_xref="SPTREMBL:O82670"		
	/translation="MRKFNLLGLMLLLCFWVCANTASVYDHTKIVIDGQRILLISG SHYRSTPEMPALFQKAKGGLDVLIQTVFNWGHPSFGKYIFEDRDLVKFIKLA QAGLVHLIRGIPYCAENWFGFPVLMKYVGIISPTDNEPFFKAMQKFTKIVSMM KAENLFQNGGIIIMSQIENEGYVWENIGAFKAYTNWAAQMGLDGVPWDMCK RFTADPIDTCNGYICENTFPNKKYPKMWTENWSGYTDFGNAICYREVEDLAYSLVA EDPONDSFVNYMYHKGTFNFGTSSGLFIATSYDAPIDEVGLTNEPKWHLRDL KATKOCEPALVSDPTITSILGNKLEAHVYSTGTSCAAFLANVDTKSAATVTEGNGKY DLPNWSYILPCKDTVFNTAKVGAGSSQKTMISTNSTDWSQSYIEEPAFSESDSIT ACALAEQIINTDSSDILWLDINSPNEKTINKGQYPLANWAGGVLHVFVNGSQL SGTVYGLVDNPLKTSNVNLTVGNKNLSLLSVAGLPNVGLHFTWNVGLGVPVTLK GLNGETRDLISWOKSYKSLGKLSLHTITGGSSVDWTOGSLAKKQPLTWYKATFN APAGNDTWHIDSSMGKIGLWSDQSIIGHWPQYIAHSGCGDCDYAGTFTNTKCRTN GNPTOTWYHPRSILNPTGNVLVLPFWGCGDSPGISLAKRI"		
CDS			

BASE COUNT	893 a	496 c	572 g	860 t
ORIGIN				

```
Query Match          47.7%; Score 1218.8; DB 8; Length 2821;
Best Local Similarity 71.9%; Pred. No. 1.7e-225;
Matches 1592; Conservative 0; Mismatches 622; Indels 0;
```

0.

Qy	23	ctaaataaaaaaacattcatttttttgaaatgtgaaaaaaatcgtacaggactaatgtgt	82
Db	163	CAAAACACACAAACATGTTGAACATATTCAAGAGATTTACAAATGAGAAAAGAAATTTAACT	222
Qy	83	tgttgattagttattgtttatttgatttttttttttttttcagtgaaagctagttttttt	142
Db	223	TGCTTGGGTGATGTTTATTACATGATGTTTGGGTGTTGCTGTTACAGCCTCAGTGACTT	282
Qy	143	atgatcacagagcataataatcataaaatgggaaaaagaaaaattctatttcttggttcaattc	202
Db	283	ATGATCACAAAACCAATGTTGATTGATGGGCACAGAGAATAATTGTATCTCTGGTCCATTC	342
Qy	203	attatccaagaagcactccacagatgtagcctgattcttatatacaaaaaggtcaaatgagag	262
Db	343	ACTATCCAAGAAGACACACCCGAGATGTGCCAGCCCTTATTTCAAAGGCCAAAAGAGAG	402
Qy	263	gcttagatgttatgaaacttatgttttctgaaatggacatgagccttctcctgaaaaat	322
Db	403	GGCTTGATGTTATCAAACTATGATATTGGAATGGACACGAGCCTTCTCTCGAAAAAT	462
Qy	323	ataattttgaaggaagatatgatctttgttagattcatcaaaatggtacaaagacagagac	382
Db	463	ATTACTTCGAAGATAGTTTGACTTGGTAAAGTTTATCAAACTAGCACACAAGTGGCC	522
Qy	383	tttatgctaattcacgtattgccccttaagtcgtgctgaatggaactttggggattcc	442
Db	523	TTTATGTTCACTTCGCGTATGTGTCCTACGTTTGTCTGAATGGAACTTTGGGGGATTC	582
Qy	443	ctgtttggcctaaaaatgtgcctggtatggaatttagacaacaaatcagccttttaagg	502
Db	583	CTGTTTGGCTCAAGTATGTTCCAGGTATTCTTTCAGAACACACACGACCATTCRAAG	642
Qy	503	tggctatgcaagattgtttcgagaaaatagtcacacatgatgaaatcgagaataattgtttg	562
Db	643	CGGCAATGCAAAAATTCACTACTAAGATTGTGAGCATGATGAAAGCAGAGAAATTTGTTTC	702
Qy	563	aatctcaagagagccaataaattatggccagatatacaaaatgagtatgaccacataaat	622
Db	703	AAATCAGGGTGTCCCAATTAATGTTCTCAGATTGAGAATGAGTATGGACAGTGAAGT	762
Qy	623	gggaaattggtctcctgctaaagcttatcaaaaatggcagctcaaatggcttaggtt	682
Db	763	GGAAATTTGGTGCACCTGGAAAGGCCTACACAAATTTGGCTGCTCAATGGCTGTAGTC	822
Qy	583	tgaaaactggtgtcccatggatcatgtgtaagcaagagatgctcctgactcgttgattg	742
Db	823	TAGATACTGGTGTACCTCGGGATATGTGCAAGCAAGAAGATGCCCTGTGATCTGTTATTG	882
Qy	743	atacttgtaatggcttctactgcgaagggttccgtccctaaataagccttcaaacctaaaa	802
Db	883	ACACTTGCATGGATATTATTGTGAAACTTTTACTCCCAACAAGAACTACAAACCCAAA	942
Qy	803	tgtggacagaagtatgagcggctgggtatacagaataatcggtgggtccaatctccaaagac	862
Db	943	TGTFGGACTGAGAAATTTGGTCCGGATGGTACACTGATTTTTGGTAAATGCTATATGTTATAGAC	1002
Qy	863	caagcgaagacattgcattttcagttgccaggtttgtttcagaacaaatggttcattttca	922
Db	1003	CAGTAGAAGACTTGGCATTAATTCAGTTGCTAGATAATCATACAAAATCGTGGGCTCTTGTTA	1062
Qy	923	attactacatgtatcatggaggaacaaattttggccggacatcatcaggggctttttcattg	982
Db	1063	ACTACTATTATGTACCATGGAGGAACAATAATTTGGCCGGACATCTAGTGGCCCTTTTCATTG	1122
Qy	983	caactagctacgatgatgatgctcctctctcgatgaatatggtgtgctgaatgaaccaaagt	1042

Db	1123	CCACAAGCTATGACTATGATGTCTTATGATGAATATGAGACTTACAAATGAACCCAAT	1182
Qy	1043	atgggcaactgagagacttacataaagctatcaagctatctgaaacggcgttagtttcat	1102
Db	1183	GGAGCCATTTTGAGAGACCTACATATAAGCAATTAAGCAATGTGAGCCGCTTGTGTCTG	1242
Qy	1103	catatgctgcggtgactagctcttgggaagtaacaagaggctcatgtttatagatcaaaa	1162
Db	1243	TGGATCTACAATTTACATCATTGGGAATTAACTTGAGGCACATGTTTACAGCAGGAA	1302
Qy	1163	ctggagctgtgctgctgtttttatccaactatgactctagattattcagtaaaagtccaact	1222
Db	1303	CTAGTGTATGTGCTGCTATTTCTTGCAATTTATGACAAAAATCTGCAGCAACGGTTACAT	1362
Qy	1223	ttcagaataggccatacaaatctgctccatgggtccatcagcaattcttccgactgcaaaa	1282
Db	1363	TTGGTAAATGGAATAACGACCTACCACCTTGGTCGTCAGACATCTTCTCATTTGCAAAA	1422
Qy	1283	ctgcggtttacaacactgcacaggtttaaactctcaagctcagagataaagatgacgctg	1342
Db	1423	CTGATGTTTTCAACACTGCAAGGTTGGTGACACAGACTCCAGAAAAACATGATTTCCA	1482
Qy	1343	caggtgggtggattctcttggcagtcatacaatgaagaacgcctactctgctgacagcg	1402
Db	1483	CAACACGATACATTTGATTTGGCAATCATATTTGAAGAACCCTGCCCTCTCTAGTGAAGATG	1542
Qy	1403	atacacttacagctaaacgaciatgggaacagaaaaacgtcacaaagattcatcagact	1462
Db	1543	ATTCTATCACAGCAAGACACTTTGGAGCACATTAATGTACTAGAGATTTCTTCAGATT	1602
Qy	1463	atctgtgttacatgacaaaatgaaaatatagcatcctaagaagattctctaagaacggaa	1522
Db	1603	ATTGTGGTACTTGACAGATGTCAACATTTAGTCCCTAATGAAGATTTTATAAGAATGGTC	1662
Qy	1523	agatccttatctcaactgttatgtccgctggtgcattgtctbttgcattgtttctgcaatggaa	1582
Db	1663	AGTATCCTATTCTGAAATGTAATGTCAGCAGGACATGTTCTACATGTGTTGTTTAAATGGTC	1722
Qy	1583	aactatcaggaaactgttatgttacattggaataatccaaaacttacatacagtggaacg	1642
Db	1723	AACCTTCAGGAACGTATATGGGTATTGGATTAATCCCTAAGTTAACTTTAGTAACAGTGT	1782
Qy	1643	tgaagttaagagctgggtataacaagattctctgctcagtggttccggtggtctcccg	1702
Db	1783	TGAACCTAACAGTTGGAATAACAGATTTCTTACTTATGTTGCTGTGGTCTCCCGA	1842
Qy	1703	acgttggcgtgcattatgatatacatggaatgcaggagttcttaggtccagtcacgtttgacg	1762
Db	1843	ATGTGGGCTTGCACTTTGAAACATGGAATGTTGGGGTCTTAGGCCAGTAACATTTGAAGG	1902
Qy	1763	gtctcaatgaagggtcaagaaacttggcgaaacagaaaatggtcttacaaggttgactctga	1822
Db	1903	GTCTAATAAGGGACTCGAGACTGTCTTTGGCAGAAATGTCCTTACAAAGGTTCGGATTGA	1962
Qy	1823	aaggcgaactgttaagctcttcaactcttaagtgggagttctcttgttgaatgggttcgag	1882
Db	1963	AAGGTGAATCATGTAGCCTTTCATACCATAACTGTGGTGGTTCCTCTGTTGACTGGACACAAG	2022
Qy	1883	gttcaataatggtctaaaaacgagcccttgaacttggtatacaaggctacatttaacgcgctg	1942
Db	2023	GATCATATTGGCTAAAAAACACCTTTGACATGTAATGAAGCAACCTTTTAATGCACCGG	2082
Qy	1943	gaggaataatccactagcttttagacatggcaagtatgggaaaggtcagatatgataa	2002
Db	2083	CTGGAAATGATCCATTGGGTCTGGATATGAGTAGCATGGGAAAAGGTGAATATGGGTAA	2142
Qy	2003	atggtgaaggcgttagtgcgccattggccttgatcatatgacagcaagcgactgcagcaaat	2062
Db	2143	ATGATCAAAGCATTTGGTCTCATTTGGCCTGGNATATATAGCACATGTGTGTGGTGTGATT	2202
Qy	2063	gcagttatgctggaacgttcaacgagaagaagtcggaacttaactgcggaacacactcttc	2122

TITLE Direct Submission
JOURNAL Submitted (10-SEP-1998) Chengappa S., Plant Sciences, Unilever Research, Sharnbrook, Bedford MK44 1LQ, UNITED KINGDOM
REFERENCE 2 (bases 1 to 2628)
AUTHORS Chengappa,S.
TITLE Heterologous Expression of a galactan specific ss-galactosidase results in extensive changes in plant cell wall composition and rheology

JOURNAL FEATURES Unpublished

source	Location/Qualifiers
	1. .2628
	/organism="Lupinus angustifolius"
	/db_xref="taxon:3871"
	/tissue_type="seed endosperm"
	/dev_stage="12 day germinating seed endosperm"
sig_peptide	130. .228
CDS	130. .232
	/codon_start=1
	/evidence=experimental
	/product="exo galactanase"
	/protein_id="CAA09467.1"
	/db_xref="GI:3860420"
	/translation="MFGSRIVMELSMRRNFHMLLLLEFWCYVYASVYDHKAIMI NGORRLILSGSHYPRSTQMPWPDLOKADGGDLVIETVYFVFNAGHEPSRGKYYPEDR FDVLGPIKLIVQOAGLEHLRIGRIPICAEWNFGFPVWLKYVPGKARTKWAQMAVGLD KPTFETVNTMAKELQSQGGPILISQIENEYGPVEIGAPKARTKWAQMAVGLD TGVPYWMCKQEDAPDIIDTCNGFYCNFTPNKNYKPKLWNTGWYAPFGGATPY PARDIAFSVARFTQNGSLFENYVMYHGGTFNFGRTSGLFVATSYDADPIDVGLLINE PKWGLHRELHRAIKQCESLNVSDPTVSPWPKNLEVLYKTESACAAPLANFYTDYST QVKGNGIDLPWSEISLPDCKTEVFNTAKVNPRLHRMTVPNSAFAMQSTNEEPA SSENDPVGTALMEQVGYTRSDSYLWLTLDVNIQNDIKDGKQVPLTAMSGHVLN VFENGYAGTATGSLDDPRLTFSSQVNLRVGNKKISLLSVSGLANGLYFTFWTNGV LGPVTLTGLSSGTDLSKQWSYKIGLKGESLSTLFTAGSNVEMVQGSILVAKKOPLA WKTPLTSPAGNDPLALDILGSMCKGEVWNGQSIGHRWPGNKARGNCNGVAGTGYD TKCLANGQPSQRYHYVPRSLWRSRGGNYLVLEWGGDPNGIALVVERT"
mat_peptide	229. .2319
	/product="exo galactanase"
BASE COUNT	835 a 449 c 581 g 763 t
ORIGIN	

Query Match 47.3%; Score 1207.4; DB 8; Length 2628;
 Best Local Similarity 73.5%; Pred. No. 2.7e-223;
 Matches 1570; Conservative 0; Mismatches 556; Indels 9; Gaps 2;

Qy	98	ttgtttatggatttttttcttcagtgaaagctagtgtttcttatgatcacagacta	157
Db	194	TGTTATTGTTTTTGGGGTTGTTATGTCACAGCCTCTGTTACTATTGATCATAAAGCCA	253
Qy	158	taatcataaatggaaaagaaaaattcttattctggttcaattcattccaaagaagca	217
Db	254	TTATGATTAAATGGCAGAGAAGATTTTGATCTCTGGTTCCTACTATCCAAAGAGCA	313
Qy	218	ctcacagatgtggcctgattcttatacaaaaggctaaagatggaggctgatgtattg	277
Db	314	CACCTCAGATGTGCCAGAGACTTATTCAAAGGCCAAAGATGGAGGGCTTGATGTTATAG	373
Qy	278	aaacttatgtttctggaatggacatagccttctctcggaataataaattttgaagaa	337
Db	374	AGACTTATGTGTTCTGGAATGGACATGAACCTTCTCTCGAAATAATATTATTTGAGGATA	433
Qy	338	gatatgctctgttagattcctcaaatgtgtacaaaagcaggaccttatgtcaatttac	397
Db	434	GGTTTGACCTTGTTGGGTTTCATAAAGTTGGTTACAGCAAGCTGGTCTATTGTTCATCTCA	493
Qy	398	gtattggcccttacgtctgtcgaatggaactttgggggattccctgttttggcctaaat	457
Db	494	GGATTGGCTCTTCATATGCTGAATGGAACTTTGGAGGATTTCCGTGTTGGCTCAAT	553
Qy	458	atgtgcctggtatggaatttagaacaacaacaatcagcccttttaaggtggctatgcaagat	517
Db	554	ATGTTCTCGTATTGCTTTTTCAGAACACACAAATGAGCCTTTTCAAGGGAGCAATGCAAAAT	613

QY 518 ttgttcagaaatagtcacatgatgaagtcagaaatattgtttgaaatctcaagagagac 577
Db 614 TCACTGAGAGATTGTTAAATATATGAAGACAGAGAAGTTGTTCAATCCACGGAGGTC 673
QY 578 caataattatggccagatatacaaaatagatgtagacagtagaattgggaattggctc 637
Db 674 CAATAAATCTGCTCAGATAGAGAATGAGTAGACAGAGTGGAAATGGTGCTC 733
QY 638 ctggttaagcttatcaaaatggcgagctcaaatggctgtagggttgaaaactggtctc 697
Db 734 CTGGAAGAAGCTTATACCAATGGGCTGCTCAATGGCTGAGGCTAGATCTAGTCTGCTC 793
QY 698 catggatcatgttaagcaagagatgctcctgatctcctgtagattgatacttgaatggct 757
Db 794 CATGGGTTATGCTCAAGCAAGAAGATCACCTGATCCTATTATTGATACCTGCAATGGAT 853
QY 758 tctactcgaaaggttcgctccttaataagccttacaacctaataatggagacagaat 817
Db 854 TTTACTGTGAAAACCTTCACTCCAACAGAACTACAACCCCAAAATTTGGACAGAAAATT 913
QY 818 ggaactggtgtatcacgaattcgtggtccaattcctcaaaagaccagccgaagacattg 877
Db 914 GGACTGGCTGGTACATGCTTTTGGTGGTCAACCCCTTATGACCCAGCAGAAATATAG 973
QY 878 cattttcagttgcaggtttgttcagacaatggttcattcttcaattactacatgtatc 937
Db 974 CATTTTCAGTTGCCAGATTTCATTGAGAATCGGGCTCACCTTTTAACACTATATGTATC 1033
QY 938 atggaggaacaaattttggcggagatcatcaagggtcttttcaattgcaactagctacgatt 997
Db 1034 ATGGAGGAACATACTTTGGCGGACATCGAATGGCCCTTCGTTGGCCACAAGTTATGACT 1093
QY 998 ataatgctctctcgatgaatattggttgcgaatgaatgaacaaagatggcacttgagag 1057
Db 1094 ATGATGCTCCCATTTGATTAATGAGACTCTTAATGATGAACCAAAATGGGGGCATCTGAGAG 1153
QY 1058 acttacataaagctatcaagctatctgaacggcttttagtttcatcatatgctgcggtga 1117
Db 1154 AATTACATAGAGCAATAAACAATGCGAGTGGGCTTTAGTGTGCGTGGTCCACAGTGT 1213
QY 1118 ctgactctggaagtaatacaagaggtcattttatagatcaaaactgtaggactgtgctg 1177
Db 1214 CATGGCTGGAAGAAACCTTTGAGGTAGATTGTTGATCAAGACAGAGTCT---GCCTGTGCTG 1270
QY 1178 cttttttatcaactatgactatagattcagtaaaagtcacaccttcagaataaggccat 1237
Db 1271 CATTTCTGCAATTAACACCCGACTATTCACGCGAAGTTAAATTCGGAAATGGACAAT 1330
QY 1238 acaatctgctccatggctccatcagcattcttcccgactgcaaaactgcccgtttacaaca 1297
Db 1331 ATGATCTACCACTTGGTCTATCAGTATCTTCTGACTGCAAACTGAAGTTTCAACA 1390
QY 1298 ctgcacagggttaactctcaagctcgagataaagatgacgctgcaaggtggtgattct 1357
Db 1391 CTGCAAGAGGTTAATTCCTCCGAGATTACATAGGAAATGACTCCAGTAACAGTGCAATTG 1450
QY 1358 cttgagcagtcatacaatgaagaacgcctactgctgtagcagcagctacacttacaagcta 1417
Db 1451 CTGCGCAGTATACAAATGAGAACCTGCATCATCAACGGAATGATCCCGTCACAGAT 1510
QY 1418 acggactatgggaacagaaacgtcaagagattcattcagactatctgtggtacatga 1477
Db 1511 ATGCACATATGGAGCAGGTTGGCGTGACCCGGATTTCTCCGATTATTGTTGGTACCTGA 1570
QY 1478 caaatgtaatatagatcatcaatgaagatttctaaagaacggaagatcccttatctca 1537
Db 1571 CAGATGTCAACATTGGTCCCTTAATGA-----TATAAGAGGATGGGAATGCGCTGTTCTGA 1624
QY 1538 ctgttatgctcgctggttcattgctctgctgttttctgctcaatggaaactacaggaactg 1597
Db 1625 CAGCAATGTCAGCAGGTCATGTTCTGATGTTTTCATCAATGGTTCATATGCGAGAACTG 1684
QY 1598 ttatggtacattggataatcccaaaacttacatcacagtggaacogtgaaagttaagactg 1657

Db 1685 CATATGGAGTCTAGATGATCCTAGATTAACTTTAGTCAAAGTGTGAATCTGAGAGTTG 1744
QY 1658 gtattacaagattctctctcgaagtgtttccgttgcgttctcccaagctgtgagctgaatt 1717
Db 1745 GCAATTAACAAGATTCTTTACTTTAGTGTTCCTGTTGGTCTCGCGAAATGTTGGTACTACT 1804
QY 1718 atgatacatgaaatcgaggagttcttagtccagtcacgttgagcggtctcaatgaaggt 1777
Db 1805 TTGAGACATGGAATCTGAGTGTCTTGGTCCAGTCACACTGACAGGCTTAAGTAGCGGAA 1864
QY 1778 caagaaactggggaacagaaatggtcttacaaggttggtctgaaaggogaaatcgtaa 1837
Db 1865 CATGGGATCTTTCGAAGCAAAAATGGCTTACAAGATTGGTCTGAAAAGGTGAAAAGCTTGA 1924
QY 1838 gtcttcaactctaagtggaagttctctgttgaatgggttcgaggttcactaatggtctc 1897
Db 1925 GCCPTCACTAGAACTGGGAGTAACTCTTGTGAATGGGTACAAGATCTTTAGTGGCTA 1984
QY 1898 aaaagcagccctgactgtgtacaaggtctacatgaagcgccttgaggaatgatccac 1957
Db 1985 AAAAAACAACCTTTGGCATGGTATAAGACAACCTTTTAGCGCACACCGCGCAACATCCGT 2044
QY 1958 tagctttagacatggcgaagtatgggaaaggtcagatatgataaaatggtgaaagcgctag 2017
Db 2045 TGGCTCTGGATTTAGGTAGCATGGGTAAAGGTGAAGTATGGGTAAATGGTCAAAAGCATTG 2104
QY 2018 gtgcattatggctggtatatacatagacaaagcgactcagcaaaatcgacttatctctgaa 2077
Db 2105 GACCCATTGGCCCTGGAATAAAGCTCTGCTGTAATTCGGCAATTTGTAATTTACCTGGAA 2164
QY 2078 cgttcaacgagaagaagtgcggaactaactgcggaacaccttctcagagatggtaccatg 2137
Db 2165 CTATACCGATACAAATGCTTAGCAAACTGTGCAAACTCCCAAGATGGTATCATG 2224
QY 2138 ttccacgatcgtggtgaaacaaagtgtgaaacttggtagtagtattcagaagaatgggag 2197
Db 2225 TTCTCGGTCTATGCTGAGATCGGTGTAACCTACTTGGTGTGCTAGAGAATGGGAG 2284
QY 2198 gtaatccaaacaggaattctctagtcaggagatca 2232
Db 2285 GTGATCTAATGGAATTCCTTTGGTGGAAAGAAC 2319

RESULT 11
A46218
LOCUS A46218 Sequence 1 from Patent WO9523228. DNA linear PAT 07-MAR-1997
DEFINITION A46218
ACCESSION A46218
VERSION A46218.1 GI:2300464
KEYWORDS
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 2628)
AUTHORS Chengappa,S., Hellyer,S.A., De,S.J. and Reid,J.S.
TITLE NOVEL EXO-(1,4)- beta -D GALACTANASE
JOURNAL Patent: WO 9523228-A 1 31-AUG-1995;
UNILEVER PLC (GB)
COMMENT Other publication AU 1713795 950911.
FEATURES
source: Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"
1..2628
130..2322
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02881.1"
/db_xref="GI:2300465"
/translation="MFGSRIVMESLMRRNFHVLVLLFPWCYVTASVTDHKAIMI
NGORRLISGSIHPSTPQMPDLIQAKDGLDVIETVFWNGHEPSPGKYFEDR
FDLVGRIKLVOQAGLFLHURIGRIFCAENWFGFPVWLKVPGLAFRTDNEPKFAMQ
KFTEKIVNIMKAELKFQSGGPIILSIIENIEYGPVWEIGAPGPKTKWAAQMAVGILD

TGVPWVCKQEDALDPIIDTCNGFYCNFTPNKNYKPKLWNTGWTAFGGATPVR
 PABDIAFSARFQNRSLFNYMYHGGTNGFRTSNGLFVATSYDYDAPIDEYGLLNE
 PKMHLRELHRAIKQCESALVSYPMTVPKKNLEVLHYKTESACAFLANYNDYST
 QVFNQVQDLPWPSISILPCKTEVENTAKVNSPRLHRKMTPVNSAFQMSNEEPA
 SSSENDPVTGYALWEQVTRDSSDYLWYLTQVNICNDIKDCKWPLYTAMSAGHVLN
 VFNQYAGTAYGSLDDPRLTFQSQVNLVGNKKISLLSVGLNANGVTHFETWNTGV
 LGPVTLLGSSGWDLSQKWSYKIGLKGESLSLHTEAGSNVSEWQSVLAKKOPLA
 WYKTTFSAPGNPLALDLGSMKGVEVNGQSIGRHPGNKARGNGCNGNYAGTYTD
 TKCLANGQPSQRWYHVRPSRLSGGNYLVLEWGGDPNGIALVERT"

BASE COUNT 835 a 448 c 581 g 764 t
 ORIGIN

Query Match 47.2%; Score 1205.8; DB 6; Length 2628;
 Best Local Similarity 73.5%; Pred. No. 5.4e-223;
 Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;

Qy 98 ttgtttatgtgattttttcttcagtgaaagctagtggtttcttattatgatgaagagacta 157
 Db 194 TGTATTGTTTTTGGTGTGTTATGTCACAGGCTCTGTACTTATGATCATAAAGCCA 253
 Qy 158 taatcataaatgggaagaataattcttattctgtgttcattcattatccaagaagca 217
 Db 254 TTATGATTAAATGGCAGAGAAGATTGATCTCTGTGTTCCATTCACTATCCAAGACCA 313
 Qy 218 ctccacagatgtgacctgattatatacaaaaaggctaaagatggagggcttagatttg 277
 Db 314 CACCTCAGATGGCCAGACCTTATCAAAAGGCCAAAGATGGAGGGCTTGATGTTATAG 373
 Qy 278 aaacttatgtttctgaaatgacatgagaccttctctctctggaataataattttgaagaa 337
 Db 374 AGACTTATGTTCTCGAATGGACATGAACCTTCTCTGGAAATATATTATTTTGGAGGATA 433
 Qy 338 gataatgattgttagattcaataatggtacaaagagcaggactttatgtcaatttac 397
 Db 434 GGTTTGACCTTGTGGGTTCATAAAGTTGTTTCAGCAAGCTGTCTATTGTTCACTCA 493
 Qy 398 gtattggcccttacgtgtgtgtgtaattggaactttgggggacttccctgtttggcttaaaat 457
 Db 494 GGATTGGTCCTTTCATATGTCTCAATGTGAACCTTTGGAGGATTTCCTGTGTTGGCTCAAA 553
 Qy 458 atgtgacctgtatgaaatttagaacaacaatcagaccttttaagtggtgacctatgcaaggat 517
 Db 554 ATGTTCCTGTGTTATGCTTTTCAGAACACACAAATGAGCCCTTCAAGAGGCAATGCAAAAAT 613
 Qy 518 ttgttcagaaaatagtcacacatgatgaagtcagaaaattgtttgaatctcaagaggac 577
 Db 614 TCACCTGAGAGATTGTAATAATATGAAGCAGAGAGATTGTTTCAATCCAGGGAGGTC 673
 Qy 578 caataattatggccacagatacaaaatgagtatggaccagtagaatagggaaattggtgtctc 637
 Db 674 CAATAAATCTGCTCAGATAGAGAATGAGTATGGACCAGTGGAAATGGGAAATTGGTGCTC 733
 Qy 638 ctggttaagcttatcaaaaatggcagctcaaatggctgtatggtttgaaactggtgtcc 697
 Db 734 CTGAAAAGCTTATACCAATATGGGCTCAATATGGGTGTAGGTCCTAGATACTGGTGTC 793
 Qy 698 catgtatcatgtgtaagcaagagatgctcctgtatcctgtgattgatacttgtaattggct 757
 Db 794 CATGGGTTATGTGCAAGCAAGAAGATGCACCTTGATCCCTATTATTGATACCTGCAATGGAT 853
 Qy 758 tctactgcgaagggttccgttacttaataagccttacaacaccttaaaatgtggacagaagtat 817
 Db 854 TTTTACTGTGAAAACCTTCACTCCAAACAAGAACTACAAACCCAAATTTGTGGACAGAAAAT 913
 Qy 818 ggactgctgtatcacaaaatcgttggttccaatctctcaagaccagccggaagacattg 877
 Db 914 GGACTGCTGTGTACACTGCTTTTGGTGGTGCAACCCCTTATAGACCACGACAGAATATAG 973
 Qy 878 cattttcagttgccaggtttgttcagacaatggtttcattcttcaattactacatgtatc 937
 Db 974 CATTTTCAGTTGCCAGATTTCATTCAGAAATCGCGGCTCCTCTTTAACTACTATATGATATC 1033

Qy 938 atggaggagaacaaattttggccggagacatcatcagggtcttttattcattgcaactagctacgatt 997
 Db 1034 ATGGAGGAATAACTTTTGGCCGGACATCGAATGGCCCTCTCTGTTGCCACAAGTTATGACT 1093
 Qy 998 atgatgctcctctcagatgaatatgggttgcgtgaaatgaacaaaagtatgggcaacttgagag 1057
 Db 1094 ATGATGCTCCCATGATGAATATGGACTTCTAAATGAACCAAAATATGGGGCATCTCTAGAG 1153
 Qy 1058 acttacataaagctcatcagcttatctgaaccggcttttagtttcatcatatctcgtcggtga 1117
 Db 1154 AATTACATAGACCAATAAACAATCGGAGTGGCCTTTAGTGTGGTGGATCCACAGTGT 1213
 Qy 1118 ctagtcttgggaagtaatacaagggtcctatgttttatagatcaaaaactctggagcttctgctg 1177
 Db 1214 CATGGCCTGGAAAAAACCTTGAGGTACATTTGTCAAGACAGAGTCT--GCCTGTGCTG 1270
 Qy 1178 cttttttatccaactatgactctcagatattcagtaaaagtcaactttcagaatagggccat 1237
 Db 1271 CATTTCTTGGCAAAATTAACAACCGACTATTCAACCAAGTTAAATTCGGAATGGACAAT 1330
 Qy 1238 acaatctgctccatggttcccatcagcatcttcccagactgcaaaaactcgcgtttttacaaca 1297
 Db 1331 ATGATCTACCACCTTGGTCTATCAGTATTTCTCTGACTGCAAAACTGAAGTTTTCACA 1390
 Qy 1298 ctgcacaggtttaactctcaaaagctcagacataaaagatgacgctcaggtggtgattct 1357
 Db 1391 CTGCAAAAGTTAAATTCGCCGAGATTACATAGGAAAATGACTCCACGATAAACATGCTT 1450
 Qy 1358 cttggcagctacataaatgaagaacgcctactgctgtagacagcagatacacattacageta 1417
 Db 1451 CTTGGCAGTCAATCAATGAAGAACCTGCATCATCAAGCGAAATATGATCCCTGCACAGAT 1510
 Qy 1418 acgactatgggaacagaaaaacgtcacaaagatctcatcagactatctggtgtacatga 1477
 Db 1511 ATGCACATATGGGACAGAGTTGGCGTGACCCCGGATCTCTCCGATTTATTTGTTGGTACC 1570
 Qy 1478 caaatgtaaatatagcatcatcaatgaaggattttctaaagacggaaagatccttatctca 1537
 Db 1571 CAGATGTCAACATTTGGTCTCAATGA-----TATAAAGGATGGGAAATGGCTGTCTGA 1624
 Qy 1538 ctgttatgtccgctgggtcattgtcttgatgttttctgtaattgaaatggaaacatacaagaaactg 1597
 Db 1625 CAGCAATGTCAAGCAGGTCATGTTCTGAATGTTTTCATCAATGGTCAATATGCAGGAAC 1684
 Qy 1598 ttatggttacattgataataatccaaaacttatacagctggcaacgtgaagttaagactg 1657
 Db 1695 CATATGGGAGTCTAGATGATCCTAGATTAACTTTAGTCAAAAGTGTGAATCTGAGAGTTG 1744
 Qy 1658 gtattacaagatttctctgctcagtggttccgtttggtctcccgaaactgttgctgctgcaatt 1717
 Db 1745 GCAATAACAAGATTCTTTACTTGTGTTCCGTTGGTCTCGCGAATGTTGTTACTACTACT 1804
 Qy 1718 atgatacatgaaatgcaggagttcttagtccagtcacgtttagcgggttctcaatgaagggt 1777
 Db 1805 TTGAGACATGGAATACTGGAGTGTGTTGTCAGTCACACTGACAGGTCTAAGTAGCGGAA 1864
 Qy 1778 caagaaacttggcgaacacagaaaatggtcttacaaggttggctgaaagcgcaatcgttaa 1837
 Db 1865 CATGGATCTTTCGAGCAAAAATGGTCTTACAGATTGGTCTGAAAGGTGAAAGCTTGA 1924
 Qy 1838 gtcttactccttaagtggtggtttctctgttgaatgggttgcaggttcaataatggctc 1897
 Db 1925 GCCTTCATACCTGAAAGCTGGGAGTAACCTGTTGAATGGGTACAAGGATCTTTTAGTGGCTA 1984
 Qy 1898 aaaagcagccctgacttgggtacaaagctacatttaacgcgccttggaagaaatgatccac 1957
 Db 1985 AAAAACAACTTTGGCATGGTATGAAGACAACCTTTTAGCGCACCCAGCCGCAACGATCCGT 2044
 Qy 1958 tagctttagacatgcaagatgggaaaaaggtcagatagataataatgggtgaagcgctag 2017
 Db 2045 TGGCTCTGGATTTTAGTAGCATGGGTGAAGGTGAAGTATGGGTAAATGGTCAAGACATTTG 2104
 Qy 2018 gtcgccattggcctggatatacatagcacaagggcagcactgagcaaaatgagttatgctgtaa 2077

QY 806 ggacagaagatggactggctggtatcagaataatcggtggtccaattccotcaaaagaccag 865
DB 976 GCACGAAATCTGGACTGGTATACAGAATTTGGTGGGCGAGTCTCTACTAGACCTG 1035
QY 866 ccgaagacaattcatttcagttgcaggttggctcagaacaatggttcattctcctaatt 925
DB 1036 CACAGGACTTGGCAATTTTCAGTTGCAAGGTTTATTCAAATGGTGGTCTCTTTGGCACT 1095
QY 926 actacatgatcatgagagaacaattttggccgacatcatcagggtcttttcattgcaa 985
DB 1096 ACTACATGATATCATGGAGAACAAATTTGGCCGAACCTGCTGGCGTCCCTTCATGTCAA 1155
QY 986 ctgactcagattatgatgctctctcgtatgaatatgggttggctgaatgaacccaagtatg 1045
DB 1156 CTAGCTATGACTATGATGACCTCTAGATGAATATGGACTACCAAGGAACCAAAATATA 1215
QY 1046 ggcaacttgagagacttacataaagctatataaagctatctgaacccgcttttagtttcaat 1105
DB 1216 GCCATTTGAAATATATGCACAAGGCATCAAGATGGCTGAGCCAGCTTTATTGGCCACTG 1275
QY 1106 atgctgcggtgactagctcttgaaagtaatacagagagctcatgtttatagatacaaaactg 1165
DB 1276 ATGCTGCAGTGTCAAACTTGGAAATATCAAGAGGCTCATGTATACCAGTCAAGATCTG 1335
QY 1166 gagcttgctgctctttttatccaaactatgactctagattatcagtaaaaaagtcacctttc 1225
DB 1336 G---TTGTGCTGCATTCCTTGCCAAATTTATGACACAAAATACCCCGTGAGAGTGACATTTT 1392
QY 1226 agaataggccatacaaatctgctcctcataggtccatcagcattctcccgactgcgaactg 1285
DB 1393 GGAATTAAGCAATATTAACCTTGGCCACCTTGGTCCATGAAGCATTTCTCTGATTGCAAAACTG 1452
QY 1286 ccgcttacaacactcacaggttaactctcaagctccagcataaagatgagcgcctgcaag 1345
DB 1453 AAGTATTCAACACTGCAGGGTTGGC---CAGAGCCCGCCCAACGAAGATGACAC---CAG 1506
QY 1346 gtggtggattctcttggcagtcatacaatgaagaacccctactctgctgatcagacgcgata 1405
DB 1507 TAGCTCACCTTCTTGGGAAGCTTACATTGAAGATGTTGCCACATCTCTGTATGACAAATG 1566
QY 1406 caattacagtaacgactatgggaacagaaaaacgctcacagaagattcatcagactatc 1465
DB 1567 CATTTACATCAGTTGGACTCTGTGAGCAAAATAAGTCTCAGTTGGGATAACACAGACTATT 1626
QY 1466 ttgtgtacatgacaaatgtaaatatagcatctaatgaaggatttctaaagaacggaag 1525
DB 1627 TGTGTACATGACATATCACAATAGGTCTTAATGACAAATTTTAAAGAACTGGAAGT 1686
QY 1526 atccttatctcactgtatgtgcgctggtgtcattgtctgtcatgtttctcgaatggaaaac 1585
DB 1687 ACCCCCACTCTCAAAGTCGATTGAGCTGGGCATGCCCTTGCATGTTTTCATCAATGGTCAAC 1746
QY 1586 tacaagaactgtttatggtacattggtataatccaaaacttacatacagtgagcaactga 1645
DB 1747 TATCAGGATCTGCTTATGGAACATTTGGCGTTCCCAAAATTAAGATTTAATCAGGGTGTGA 1806
QY 1646 agttaagagctggtattacaagatttctctgctcagtggttccgttggtctcccgcaag 1705
DB 1807 AGCTGAGAGCTGGCATAACAAGCTTGCTTACTTATGTTTTCAGTTGGTCTGGCGCAATG 1866
QY 1706 ttggcgtgcattatgatacattggaatgcaggaggttcttaggtccagtcagcttgagcgtc 1765
DB 1867 TTGCTCTACACTTTGAGACATGGAATAGTGGAGTCTTGGCGCGGTCCGCTTGGCTGGTGTG 1926
QY 1766 tcaatgaagggtcaagaactggcgaacagaagaatggtcttacaaggttggtctgaaag 1825
DB 1927 TCAATTGGGGAACATGGGATATGACACCATGGCAATGGCACTTACAAGATTGGCATGAGAG 1986
QY 1826 gcgaactcgttaagcttccactccttaagtggaggttctctgttgtaagggttcgaggtt 1885
DB 1987 GTGAGATATGAGCCTCCATACCGTCACTGGAAGTCTCTCTGTTGAATGGGTACAAGGAT 2046

QY 1886 cactaatggtcctaaagcagccctgacttggtatacaaggtctacatttaacgcgcctggag 1945
DB 2047 CACTATTGGCTCAATATCGACCCCTTACATGTTATAGGCTATTTTAAATGCACCACCAG 2106
QY 1946 gaaatgatccactcatttttagacatggcaagtatagggaaaaaggtcagatatgataaaatg 2005
DB 2107 GAAATGCACCGTGTAGCTTTAGATATGGGTAGCATGGGAAAGGTCAAATGTGATTAAATG 2166
QY 2006 gtgaaggcgtaggtcgccattggcctgatacatgatacagaagcgactgcagcaaatgca 2065
DB 2167 GTCAGAGCATGGACGCCCATTTGCCGGCATATAAAGACACACGGGAGTTGTGGTGTGTT 2226
QY 2066 gttatgtggaaacttcaacgagaagaagtgcggaactaaactgcggaacacctctcaga 2125
DB 2227 ACTATGCTGGTACTTATACTGAAACAAATGTCGAACAAATTTGTGCAGACCCCTCTCAGA 2286
QY 2126 gatgtacatggttccacgactggtgctgaaacaaagtggaaaacttggtagtagtattcg 2185
DB 2287 GATGTCACATGTTTCCCGCTCGTGTGTTGAAGTCAAGTGGGAATTTATTGGTGTGTTG 2346
QY 2186 aagaatgggaggttaatccaacaggaatttctctctagtcaggaga 2229
DB 2347 AGAATGGGGTGGTATCCGACTAAGATTTCTTTCGTAGCAAGA 2390
RESULT 14
AF004812
LOCUS AF004812 2202 bp mRNA linear PLN 22-JUN-1997
DEFINITION Mangifera indica beta-D-galactosidase (SP26) mRNA, partial cds.
ACCESSION AF004812
VERSION AF004812.1 GI:2209357
KEYWORDS
SOURCE mango.
ORGANISM Mangifera indica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosidae; eucotyledons II; Sapindales; Anacardiaceae; Mangifera;
Paria-Arenas,S., Cruz-Hernandez,A., Gutierrez-Martinez,P. and
Gomez-Lim,M.A.
Cell wall metabolism in mango fruit: isolation of a cDNA clone to
beta-D-galactosidase and expression studies
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2202)
AUTHORS Paria-Arenas,S., Cruz-Hernandez,A., Gutierrez-Martinez,P. and
Gomez-Lim,M.A.
Direct Submission
JOURNAL Submitted (19-MAY-1997) Genetic Engineering, CINVESTAV-Trapuato,
Apartado Postal 629, Irapuato, GTO 36500, Mexico
FEATURES
source
1..2202
/organism="Mangifera indica"
/strain="Vanilla"
/db_xref="taxon:29780"
/tissue_type="fruit"
<1..>2202
/gene="SP26"
214..>2202
/gene="SP26"
/EC_number="3.2.1.23"
/function="cell wall degradation; hydrolysis of
(1-4)-beta-D-galactans"
/codon_start=1
/product="beta-D-galactosidase"
/protein_id="AAB61470.1"
/db_xref="GI:2209358"
/translation="MGPEGSMTDLITNMFLLMLFSSWCFVEATVSDVHKAIID
GQRRLISGSIHYPRSTPQMKDQVDVQTVFVNGHEPSPGKYFFDRDY
LVRFILVQOAGLYVHLRIGPVCAENFGFPVWLKYVPGIEPTNEPFFKAAMQKF
TEKIVSNMKAELFLFETGGPIILSOIENEFPGVWEIGAPCKAYTKWAQMAVGLDTG
VPVMCKQDDADPDVINTCNFGYCFENFVNQKNPKMWTENWGNWTFAGGPTQRP
EDVAFSVARITQNGGSPVNTYMTHTGNTFGRTAGGPIATSYDIDAPLEGLUREPK
WGLRDLJHKAICLESALVSTDPVTVSLGNNQNEVHFNPSPKSGSCAAFLANYDTTSSAK

VNFKIMQYELPPWSISILPDKCTAVFNARLGAOSSLKQMPVSTFSWQSYIESASS
SDDKTFYTDGWEQINVTDRASDYKLVNINISNEGLKNGQDPLTITWSAGHALH
VFINGLSGTGYGVNDPKLTFSONKVRVGNVNDLSLSISVGLQNVTHFEQWNTGV
LGPVTLRLNBSERDLKQOQWSYKIGLGEDLSLHTVSGSSSVVEBSSLAQKPLT
WYKTFNPAGNEPLALDMSTMWKGILWINSQSIGRPRP"
BASE COUNT 620 a 431 c 491 g 659 t 1 others
ORIGIN

Query Match 46.3%; Score 1181.6; DB 8; Length 2202;
Best Local Similarity 75.7%; Pred. No. 2.6e-218;
Matches 1489; Conservative 1; Mismatches 470; Indels 6; Gaps 2;
QY 62 aaatgtaaggactaatgtgtgtgtattattattgttttttttttttttttttttttt 121
DB ACATGTTGATACCAACCATGTTTATGTTGCTGCTAATGCTGTTTCTTCATCGGTTTGT 298
QY 122 caagtgaagctagtgtttcttatgatgacagagctataataataatgggaaaaa 181
DB TTGTGGAAGCAACTGTGCTATGACCAATAAAGCTATCATCTATGATGCGCCAGAGA 358
QY 182 tcttatcttggttcaattcattatccaagaagcactccacagatgtggcctgactta 241
DB TTCTTATTTCTGGCTCCATTCATTAATCCAGAGAGTACTCTCAGATGTGGCCTGATCTTA 418
QY 242 taaaaaggctaaagatggagccttagatgtttatgaaacttatgttttcttggaaatggac 301
DB TTTCAGAGGCCAAAGATGGAG--TTGATGTTATACAACTTATGCTCTCGAATGGAC 475
QY 302 atgagccttctcctggaaaaataattttgaaaggaatgatattgtttagattcata 361
DB ATGAGCCCTTCTCTGGGAAATATTACTTTAGGACAGATATGATCTGTTCTGGTTCAATTA 535
QY 362 aaatgtaacaaagcaggactttatgtcaatttcaatttgaattgtggcctcactgtgctg 421
DB ACCTGTTTCAGCAAGCTGGCCCTTTATGTTTCATCTCCGGANTGGCCCTATGTTTGTGACG 595
QY 422 aatggaactttggggattccctgttttgctaaaaataatgtcgtggtatggaaatttagaa 481
DB AATGGAACCTTTGGGGGTTTCTCTGTTTGGCTGAATATGTTCTCGAATTTGAATTTAGNA 655
QY 482 caaacaatcaagccttttaagtggtgctatgcaaggatttgttcagaaaaatagtcacatga 541
DB CAGACAATGAACCTTTCAAGGGCGGAATGCAAAAATTCACGGAGAGATAGTCAGCATGA 715
QY 542 tgaagtcagaaaattgtttgaactcgaagagaccataattatgtgccagatacaca 601
DB TGAAGCGAGAAAGTTGTTGAACACAGAGTGGTGGTCCCAATAATTTTGTCCACAGATAGA 775
QY 602 atgagatggaccagtagaattgggaaattggtgctcctggtaaagcttatatacaaaatggg 661
DB ATGAATTTGGACCCCTTGAATGGGAAATTTGGCGCTCTCTGGTAAAGCTTATACGAAATGGG 835
QY 662 caagctcaaatgcttaggtttgaaactggtgtcccatggtatcatgtgtaagcaagagg 721
DB CTGCTCAAAATGGCTTGTGCTGTGACACTGGTGTGCCCATGGTATGTTGTCGAAACAGATG 895
QY 722 atgctccctgactcctgtgattgatactgttaattggtccttactgcgaagggttccctccta 781
DB ATGCTCTCTGATCCAGTTATAACACTTGAATGGTTTCTACGTGAAAAATTTTGTGCCCA 955
QY 782 ataaagccttacaaactaaaaatgtgacagaagtagtgactggtggttatatacgaattgg 841
DB ACCAGAAAAATAACCCAAAAATGTGACAGAAAACTGGACTGGCTGGTTCACTGCGTTTG 1015
QY 842 gtggtccaattcctcaagaccagccgaagacattgtcatttcaagttgccaagttgttc 901
DB GTGGTCCAACTCCTCAAGACCAGAGAGAGCGTGGCTTTTCTGTGGCAAGTTTCATAC 1075
QY 902 aqaacaatgtttcattcttcaattactacattatcatatgagagacaaattttggccgga 961
DB AGAATGGAGGTTTCATTCGTGTAATCTACTACATGATACCATGGAGAACTAATTTTGGCCGGA 1135

QY 962 catcatcagggtcttttcattgccaactagctacgattatgatctcctctcgtatgaatg 1021
DB CAGCGGTGGCCCTTCATTGCTACTAGTATGATTATGATGCTCTCTTTAGATGAATAG 1195
QY 1022 gtttctgtaatgaacaaaagtattgggcaacttgagagaaacttacaaagctatacagctat 1081
DB GATTACTAAGGGAACCAAAATGGGGCATTTTGAGAGATTTGATTAAGGCATCAACATAT 1255
QY 1082 ctgaaccggcttttagtttcatcatatgctgcggtgactagttcttggaagtaactcaagagg 1141
DB GTGAATACAGCTTTAGTTTCGACAGATCCACACTGTCACCTCGGAAATATCAAGARG 1315
QY 1142 ctcatgtttatagatcaaaactggagccttgctgctgttttttttccaaactatgactcta 1201
DB TTCAATGATTTAAACCCAAAATCTGGATCTGTGCGGCATTCCTGCTAACTATGACACAA 1375
QY 1202 gatattcagtaaaagtcacctttcagaataggaatagccatacaaatctgcctccatagttcata 1261
DB CATCTCGGCCAAAAGTGAAATTTCAAATAATGCAATATGAACTGCCACCTTGGTCCATCA 1435
QY 1262 gcaattctccgactgcaaaactgcgctttcaaacactgcacaggtttaactctcaagct 1321
DB GCATCTCTCTGACTGCAAAACTGCAGTTTAAACACTGCTAGGCTTGGTGGCCAAAGCT 1495
QY 1322 cgagcataaagatgacgcctgcaggtgtggattctctctggcagtcatacaaatgaagaaa 1381
DB CCTGAAACAGATGACACCT--GTCAGCACATTTTCTTGGCAGTCAATACATTTGAAGAA 1552
QY 1382 cgcctactgtgatagacagcgatacacttacagctaaacgactatgggaacagaaaaacg 1441
DB GTGCGCTCTTCTAGTGATGATAAGACTTTTACCACCTGATGGGCTGTGGGAACAATTTAAATG 1612
QY 1442 tcacaagattcatcagaactatctgtgtacatcaacaaatgataatagcatcatag 1501
DB TGACTAGAGATGCTTCGGACTACTTGTGTATATGACAAATATTAATATAGACTCTAACG 1672
QY 1502 aaggattttcaagaacgaaaggatccttatctcactgttatgtcgcgtgctcatgtct 1561
DB AAGGATTTCTGAAGAATGGACAAGTCTCTTCTTACCAATTTGGTCTGCTGCTATGCAC 1732
QY 1562 tgcagtgtttcgtcaatggaaaaactatcaggaactgttttatggtaacttggataatccaa 1621
DB TGCATGTTTTCATTAATGGTTCAGCTATCAGGAACGGTTTATGGTGGAGTTGCAATCCAA 1792
QY 1622 aactcacacagtggaacgtgaagttaagagctgggtatttaacaagattctctgctca 1681
DB AATTAACTTTCACTCAAAATGTCAAAATGAGAGTTGGTGTTAACCAAGCTTTCTTTATTA 1852
QY 1682 gtgttccggttggtctcccgaacgttggtgctgattatgatatacatgagagagttc 1741
DB GTATATCTGTGGTCTTCAGAAATGTCGGCACACATTTTTCAGCAATGGAATACTGGGGTGC 1912
QY 1742 taggtccagtcacgttgagcgtctcaatgaagggtcaagaaacttggcgaacagaaat 1801
DB TAGGTCAGTAACATTTAGGGGTCTCAATGAGGGGACGAGACTTGTCTAAACAGCAAT 1972
QY 1802 ggtctacaaggttggtctgaaagcgcaatcgttaagtcttctcactccttaagtgaggagt 1861
DB GGTCTTACAAGATTGGTCTAAAAGGTGAGGATTTAAGCCTTCTATCTGTGTAGTGGGAGCT 2032
QY 1862 ctctgttgaattggttcgaggttccactaatggctcaaaagcagcccttgaacttggta 1921
DB CCTCTGTTGAATGGGTAGAAGGATCATCTTTGGCCCCAAAAACAACCATTTGACATGTTG 2092
QY 1922 aggtcacatttaacgcgctggagaaaatgatccactagcttttagacatggcaagtatgg 1981
DB AGACTACTTTCAATGCACCGGGGCAATGAACCTTTAGCTTTTGGATATAGTACCATGG 2152
QY 1982 gaaaaggctcagatggataaaatggtgaaggcgttaggtcgccattg 2027
DB GAAAAGGGCTGATTTGGATAAAACAGCCAGAGCATTTGGGCGCCCTCG 2198

Db 1197 GTGTTTGTGCAGATATAAACGAATGGCAACCACAGTTGGTGAATTGGC 2321

1197 AGCTGGCTCTTGTGTGCAATCTTGTAACTACGACCAACACTCTCTTTTGTACTGTGT 1256
1220 cctttcagataggccatataaatctgctccatggttcacagcattcttccgactgca 1279
1257 CATTTGCAACAGCATTACAACTTGGCCACCATGGTCAATCAGCANCTCTTCCCGACTGCA 1316
1280 aaactgcggtttacaacactgcagaggttaactctcaagctcgagcataaaagatgacgc 1339
1317 AGAACACTGTATTTAATACAGCAGGATCGGTCTCAAAAGTGTCAAGTCAAGTCAAGTCACTC 1376
1340 ctgcaggtggtgattctcttggcagtcatacaataaagaacagcctactgctgatgaca 1399
1377 CAGTCAGCAGAGATTGCCCTGGCATCTAATGAAGACATCATCTTATCAAGACA 1436
1400 gcatcacacttacagctaacgagctatgggaacagaaaaacgtcacaagagattcatcag 1459
1437 G---TAGTTTACAGTGTGGCTATTGGACAGATATAATACAAAGAGAGCTGTCTG 1493
1460 actatctgtgtacatgacaaatgtaaatatagcatctaatgaaggattttctaaagaacg 1519
1494 ATTATTGTGGTATTCAACAGATGTCAAGATTGATTCAAGAGAAAAGTTTTCAGAGGCG 1553
1520 gaaagatccttatctcactggttatccgctggtcatgctcttgcatggtttttogtcaatg 1579
1554 GAAATGGGCTTGGCTTACGATCATGTCACTGGGATGCAATGCATGTTTGTGTAATG 1613
1580 gaaactatcagaactgtttatgttacattggtataatccaaacattacatacagtgga 1639
1614 GTCAAATTAGCAGAACTGCNATGGAAAGTTAGAAAACCGAAACCTAACTTCAGTAAAG 1673
1640 acgtgaagttgaagctggtattacaagaattctctgctcagtggttccgttgggtctcc 1699
1674 CCGTAAATCTGACGAGCAGGTGTTAACAAGATTCTCTACTGAGCATTCGCTGGGCTTC 1733
1700 cgaactgtggcgtgattatgatacatggaatgcagagagttcttagtccagtcacgttga 1759
1734 CGAATATCGGCCACATTTTGACATATGGAATGCTGGTGTCTTGGGCCAGTCTCACTAA 1793
1760 gcggtctcaatgaagggtcaagaaacttggcgaacagaaaaatggtctctacaaggttggct 1819
1794 CTGGCTCTTGACGAGGGGAAAGAGATTAACTGGCAGAAATGGTCTTACAAAGTTGGTC 1853
1820 tgaagggcaactcgttaagtcttcaactccttaagtgggagttctctctgttgaatgggttc 1879
1854 TAAAGGAGAAGCCTTGAGCCTCCATTCACTCAGTGTGTAGCTGCTCAGTTGAGTGGGTGC 1913
1880 gaggttcactaatggttcaaaaagcagccctgacttgggtacaaggctacatttaacgcgc 1939
1914 AGGGTTCTTTAGTGGCTCAGAGACGCCACTACATGGTACAAGAGCACTTTTAAATGCTC 1973
1940 ctggaggaaaatgatccactagcttttagacatggaagtagtggaagagtcagatatgga 1999
1974 CAGCTGGAATATCTCTTGGCTTTAGACTTGAATACCACTGGGCAAGAGCAAGTGTGGA 2033
2000 taaatggtgaagcgttagtgcgcttggcctgtgatacatagcacaaggcgactgcagca 2059
2034 TAAATGGTCAAAAGCCTCGGAGCGTATTGGCTGGATATAAAGCATCTGTAACTGCGGTG 2093
2060 aatgcagttatgctgaaagcttcaacgagaagtcgaggaaactaaactcgggacaacctt 2119
2094 CCTGTAACTATGAGGCTGGTGTAAATGAGAAAAAATGCCTAAGTAACTGTGGAGAGCCTT 2153
2120 ctgagatggttaccatgctccacgactgctgctgaaacaaagtgaactgtttagtag 2179
2154 CACAACGATGGTATCATGTTCCCGTTCCTGGCTGTATCTCTACTGGAATTTGTTAGTTC 2213
2180 tattcgaagaatgggaggttaatccacaggaatttctctagtcaggagatcaagataaa 2239
2214 TATTTGAGGAATGGGAGGAGAGCCCTCATGGAATCTCTTTGGTAAAAAGAGAAGTTGCAA 2273
2240 gaactcgaagaagtaaaactgttcaactatggtgcttgaattcgc 2287

Search completed: June 25, 2002, 06:21:18
Job time: 13608 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:37:40 ; Search time 34.31 Seconds
(without alignments)
3650.490 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLLLVICLLDFSS.....VVFEWGGNPTGISLVRRSR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPATREMBL09**
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviris:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	3924	100.0	724	10	081100	081100 lycopersico	
2	3909	99.6	724	10	Q9T0P6	Q9t0p6 lycopersico	
3	3632	92.6	724	10	Q93XM4	Q93xm4 capsicum an	
4	3063.5	78.1	721	10	Q92P30	Q92p30 carica papa	
5	3055	77.9	731	10	Q9AYS1	Q9ays1 pyrus pyrif	
6	3028.5	77.2	721	10	Q9M5J4	Q9m5j4 phaseolus a	
7	2992.5	76.3	730	10	Q65736	Q65736 cicer ariet	
8	2948.5	75.1	723	10	Q82670	Q82670 cicer ariet	
9	2941	74.9	838	10	Q9ZP11	Q9zpl1 lycopersico	
10	2928.5	74.6	847	10	Q9SCW1	Q9scw1 arabidopsis	
11	2915	74.3	843	10	Q93X58	Q93x58 fragaria an	
12	2906	74.1	730	10	Q9ZP17	Q9zpl7 lupinus ang	
13	2901.5	73.9	722	10	Q93X56	Q93x56 fragaria an	
14	2848.5	72.6	728	10	Q9SCV0	Q9scv0 arabidopsis	
15	2838	72.3	729	10	Q9SZ15	Q9sz15 arabidopsis	
16	2809.5	71.6	724	10	Q93Y27	Q93y27 arabidopsis	

17	2808.5	71.6	724	10	Q9SCV8	Q9scv8 arabidopsis
18	2807	71.5	727	10	Q9SCW0	Q9scw0 arabidopsis
19	2797	71.3	727	10	Q9LPA6	Q9lfa6 arabidopsis
20	2736	69.7	663	10	Q04976	Q04976 mangifera l
21	2687.5	68.5	854	10	Q94B17	Q94b17 vitis vinif
22	2653.5	67.6	745	10	Q92RV9	Q92rv9 cicer ariet
23	2636.5	67.2	845	10	Q9LIS9	Q9lls9 lycopersico
24	2617.5	66.7	739	10	Q9M5J3	Q9m5j3 phaseolus a
25	2608	66.5	853	10	Q23243	Q23243 arabidopsis
26	2608	66.5	856	10	Q9SCV9	Q9scv9 arabidopsis
27	2572	65.5	853	10	Q42150	Q42150 arabidopsis
28	2560	65.2	732	10	Q9MAJ7	Q9maj7 arabidopsis
29	2560	65.2	732	10	Q9SCV7	Q9scv7 arabidopsis
30	2556	65.1	732	10	Q93263	Q93263 arabidopsis
31	2369.5	60.4	707	10	Q65761	Q65761 cicer ariet
32	2261	57.6	852	10	Q9SCV4	Q9scv4 arabidopsis
33	2259.5	57.6	839	10	Q9SK11	Q9sk11 arabidopsis
34	2219	56.5	852	10	Q9LLT0	Q9llt0 lycopersico
35	2177	55.5	840	10	Q93X57	Q93x57 fragaria an
36	2103.5	53.6	741	10	Q9FN08	Q9fn08 arabidopsis
37	2097.5	53.5	741	10	Q9SCV2	Q9scv2 arabidopsis
38	2049.5	52.2	887	10	Q9SCV3	Q9scv3 arabidopsis
39	2035.5	51.9	895	10	Q48836	Q48836 arabidopsis
40	1843	47.0	788	10	Q9SCV5	Q9scv5 arabidopsis
41	1811	46.2	718	10	Q9SCV6	Q9scv6 arabidopsis
42	1808	46.1	718	10	Q9FFN4	Q9ffn4 arabidopsis
43	1786	45.5	779	10	Q9C6W4	Q9c6w4 arabidopsis
44	1682	42.9	715	10	Q9FSF9	Q9fsf9 nicotiana t
45	1649	42.0	808	10	Q94728	Q94728 oryza sativ

ALIGNMENTS

RESULT 1
ID 081100 PRELIMINARY; PRT; 724 AA.
AC DT 01-NOV-1998 (TREMBlrel. 08, Created)
AT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN BGAL4.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RUTGERS; TISSUE=Tomato FRUIT;
RX MEDLINE=98289087; PubMed=9625694;
RA Smith D.L., Starrett D.A., Gross K.C.;
RT "A gene coding for tomato fruit beta-galactosidase II is expressed during fruit ripening. Cloning, characterization, and expression pattern."
RT Plant Physiol. 117:417-423(1998).
RL -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC EMBL; AF020330; AAC25984.1; -
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF013101; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRLASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase; Transil peptide.
FT TRANSIT 1 23 POTENTIAL.
FT ACT_SITE 181 181 PROTON DONOR (BY SIMILARITY).
SQ SEQUENCE 724 AA; 80513 MW; 6F369499E25BDB2C CRC64;

Query Match 100.0%; Score 3924; DB 10; Length 724;
Best Local Similarity 100.0%; Pred. No. 1.le-256;


```
Q93XM4
ID Q93XM4 PRELIMINARY; PRT; 724 AA.
AC Q93XM4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE.
GN BGI.
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MC11; TISSUE=RIPENING FRUIT PERICARP;
RA Zainal-Abidin W., Zainal Z., Ali Z.M., Lazan H.;
RT "A full length cDNA sequence of Capsicum annuum L. beta
galactosidase.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY029226; AAK40304.1; -.
SQ SEQUENCE 724 AA; 80176 MW; DE9BD5B6E09D4DC9 CRC64;

Query Match 92.6%; Score 3632; DB 10; Length 724;
Best Local Similarity 91.2%; Pred. No. 5.7e-237;
Matches 662; Conservative 38; Mismatches 22; Indels 4; Gaps 2;

QY 1 MLRTN--VLLLVICILDFSSVKASVSYDDRAIIINGKRKILSGSIHYPRSTPQWMPD 58
DB 1 MKMSNNVLLVVLVTCSDL--VKANSYDDRAIVNGKRKILSGSIHYPRSTPQWMPD 58
QY 59 LIQAKDGGDLVITYVFWNGHEPSPKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYC 118
DB 59 LIEAKDGGDLVITYVFWNGHEPSPKYNFEGRYDLVRFIKLVQAGLYVNLRIQPVYC 118
QY 119 AENFNGFPVWLKYVPGMEFTNNQPKFVAMQGVQKIVNMKSENLFESQGGPIIAQI 178
DB 119 AENFNGGLPVWLKYVSGMEFTNDQPKFVAMQGVQKIVNMKSEKLFEPQGGPIIAQI 178
QY 179 ENYGVPEWELGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCGFR 238
DB 179 ENYGVPEWELGAPKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCGFR 238
QY 239 PNPKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNG 298
DB 239 PNPKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNG 298
QY 299 RTSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSG 358
DB 299 RTSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYPTVTSLSG 358
QY 359 EAHVYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQNVSQ 418
DB 359 EAHVYRSKSGACAAFLSNYDAKYSVRVSNQNLPYDLPPWISILPDCKTAVYNTAKVSQ 418
QY 419 SSSIKMTPAGGLSWQSYNETPTADSDTLTANGLMEQKNVTRDSDYLWYMTNVNIAS 478
DB 419 GSSIKMTPAGGLSWQSYNETPTADSDTLRANGLMEQKNVTRDSDYLWYMTDVNIAS 478
QY 479 NEGLKNGKDPYLVMSAGHVLHVFNKGLSGTYGTLNDNPKLYSGNVKLKLAGINKISL 538
DB 479 NEGLKNGKDPYLVMSAGHVLHVFNKGLSGTYGTLNDNPKLYSGNVKLKLAGINKISL 538
QY 539 LSVSVGLPNVGVHVDTNAGVLGVPVTLISLNGEGRNLAKQKWSYKVGKLGESLSLSLG 598
DB 539 LSVSVGLPNVGVHVDTNAGVLGVPVTLISLNGEGRNLAKQKWSYKVGKLGESLSLHLSG 598
QY 599 SSSVWVVRGSLMAOKPLTWYKATFNAPGGNDPLALDMSMGKGOIWTNGSVGRHWPY 658
DB 599 SSSVWVVRGSLMAOKPLTWYKATFNAPGGNDPLALDMSMGKGOIWTNGSVGRHWPY 658
QY 659 IAOGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKP SGNLLVVFEEWGGNPTGIS 718
DB 659 IAOGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKP SGNLLVVFEEWGGNPTGIS 718

659 AAOGDCSKCSYAGTFNEKKCTQNCQPSQRWYHVPWSLKP SGNLLVVFEEWGGNPTGIS 718
719 LVRRSR 724
719 LVRRSR 724

RESULT 4
ID Q92P30 PRELIMINARY; PRT; 721 AA.
AC Q92P30;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
OS Carica papaya (Papaya).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Caricaceae; Carica.
OX NCBI_TaxID=3649;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. EKSOTIKA; TISSUE=MESOCARP;
RA Othman R., Choo T.S., Ali Z.M., Zainal Z., Lazan H.;
RT "A full-length beta-galactosidase cDNA sequence from ripening
papaya.";
RL Plant Physiol. 118:1102-1102(1998).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC -1- GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
DR EMBL; AF064786; AAC77377.1; -.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 721 BETA GALACTOSIDASE.
SQ SEQUENCE 721 AA; 80985 MW; BA1BF5E16C299617 CRC64;

Query Match 78.1%; Score 3063.5; DB 10; Length 721;
Best Local Similarity 75.2%; Pred. No. 1.4e-198;
Matches 543; Conservative 86; Mismatches 90; Indels 3; Gaps 2;

QY 1 MLRTNVLVLLVLCILDFSSVKASVSYDDRAIIINGKRKILSGSIHYPRSTPQWMPDLI 60
DB 1 MLRTNVLVLLVLCILDFSSVKASVSYDDRAIIINGKRKILSGSIHYPRSTPQWMPDLI 60
QY 61 QKAKDGGDLVITYVFWNGHEPSPKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAE 120
DB 61 QKAKDGGDLVITYVFWNGHEPSPKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAE 120
QY 59 QNAKEGGLDVLTQTVFWNGHEPSPKYNFEGRYDLVRFIKLVHQAGLYVHLRISPYICE 118
DB 59 QNAKEGGLDVLTQTVFWNGHEPSPKYNFEGRYDLVRFIKLVHQAGLYVHLRISPYICE 118
QY 121 WNFEGFPVWLKYVPGMEFTNNQPKFVAMQGVQKIVNMKSENLFESQGGPIIAQIN 180
DB 121 WNFEGFPVWLKYVPGMEFTNNQPKFVAMQGVQKIVNMKSENLFESQGGPIIAQIN 180
QY 119 WNFEGFPVWLKYVPGIOPRTDNGPPKFAQMQKFTKIVNMKAEKLFEPQGGPIIMSQJEN 178
DB 119 WNFEGFPVWLKYVPGIOPRTDNGPPKFAQMQKFTKIVNMKAEKLFEPQGGPIIMSQJEN 178
QY 181 EYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCGFRPN 240
DB 181 EYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCGFRPN 240
QY 179 EYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCENFMFN 238
DB 179 EYGPVWEIGAPGKAYTKWAAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYCENFMFN 238
QY 241 KPYKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNGRT 300
DB 241 KPYKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNGRT 300
QY 239 ANYKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNGRT 298
DB 239 ANYKPKMTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNGSFENFYMYHGGTFNGRT 298
QY 301 SSGFLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSG 360
DB 301 SSGFLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSG 360
QY 299 AGGFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSG 358
DB 299 AGGFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSG 358
QY 361 HVIYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQNVSQ 420
DB 361 HVIYRSKSGACAAFLSNYDSRISVKYTFQNPYNLPWISILPDCKTAVYNTAQNVSQ 420
QY 359 HVFWTKT--SCAAFLANLYDLKYSVRVTFQNLPPWISILPDCKTAVYNTAQNVSQ 417
DB 359 HVFWTKT--SCAAFLANLYDLKYSVRVTFQNLPPWISILPDCKTAVYNTAQNVSQ 417
```



```
Db 10 VLMLFFWCVGTVASYVDYDKAIVIDGKRILLISGSIHYPRSTPQMPDIIQAKDGLD 69
QY 70 VIETVFWNGHEPSPGKYNEFGYDLVRFIKWQVAGLYNLRIGPVVCAEWNFGGPPVW 129
Db 70 VIQTVFWNGHEPSPGKYNEFGYDLVRFIKWQVAGLYNLRIGPVVCAEWNFGGPPVW 129
QY 130 LKVPYVMEFRTNQPFKAMQGVQKIVNMKSENLFESQGGPIIQAQIENIEYGPVWEI 189
Db 130 LKVPYVMEFRTNQPFKAMQGVQKIVNMKSENLFESQGGPIIQAQIENIEYGPVWEI 189
QY 190 GAFKATYKWAQMAVGLKTVGVPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKPKMT 249
Db 190 GAFKATYKWAQMAVGLKTVGVPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPKPKMT 249
QY 250 EVMTGWTKFGGPIIPQPAEDIAFSAVAREVQNGSFYNYMHGCTNFGTSGLFIATS 309
Db 250 EVMTGWTKFGGPIIPQPAEDIAFSAVAREVQNGSFYNYMHGCTNFGTSGLFIATS 309
QY 310 YDYPADLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEAHVYRSKGA 369
Db 310 YDYPADLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEAHVYRSKGA 369
QY 370 CAAPLSNYSRYSVKVTFQRPYNLPWISILPDKCTAVNTAQNVSQSSIKMTAGG 429
Db 370 CAAPLSNYSRYSVKVTFQRPYNLPWISILPDKCTAVNTAQNVSQSSIKMTAGG 429
QY 430 GLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLVMTNVTNATSEGLFKNGKDP 489
Db 430 GLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLVMTNVTNATSEGLFKNGKDP 489
QY 490 YLVMSAGHVLHVFVNGKLSGTYGTLDNPKLTVSGNVKLAGINKISLSVSVGLPNVG 549
Db 490 YLVMSAGHVLHVFVNGKLSGTYGTLDNPKLTVSGNVKLAGINKISLSVSVGLPNVG 549
QY 550 VHDYTNAGVLGPTVTLGSLNEGSRLAKQKWSYKVLGKESLSLHSLSGSSVVEWVRGSL 609
Db 550 VHDYTNAGVLGPTVTLGSLNEGSRLAKQKWSYKVLGKESLSLHSLSGSSVVEWVRGSL 609
QY 610 MAQKQPLTWKATFNAPGNDPLALDMASMGKGQIWTNGEGVGRHWPYIAQGCSCSY 669
Db 610 MAQKQPLTWKATFNAPGNDPLALDMASMGKGQIWTNGEGVGRHWPYIAQGCSCSY 669
QY 670 AGTFNEKKQTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 723
Db 670 AGTFNEKKQTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 723
QY 671 AGTYDQKCRKTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 723
Db 671 AGTYDQKCRKTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 723
RESULT 7
ID 065736 PRELIMINARY; PRT; 730 AA.
AC 065736;
[1]
DT 01-NOV-1998 (Tremblrel. 07, Created)
DE 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.
RC SPRAIN-CV, CASTELLANA; TISSUE-ETIOLATED EPICOTYLS;
RA Esteban R., Dopico B., Labrador E.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -/- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -/- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL: AJ005042; Cicer.
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
```

```
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 730 AA; 81300 MW; F79252ABFBD1E6C3 CRC64;

Query Match 76.3%; Score 2992.5; DB 10; Length 730;
Best Local Similarity 73.9%; Pred. No. 8.8e-194;
Matches 530; Conservative 88; Mismatches 96; Indels 3; Gaps 2;

QY 8 LLLVICLLDFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDIIQAKDGG 67
Db 16 LVLELCL--FVSVTASVYVDHKAIVINGQRILLISGSIHYPRSTPQMPDIIQAKDGG 73
QY 68 LDVIETVFWNGHEPSPGKYNEFGYDLVRFIKWQVAGLYNLRIGPVVCAEWNFGGPP 127
Db 74 VDVITQTVFWNGHEPSPGKYNEFGYDLVRFIKWQVAGLYNLRIGPVVCAEWNFGGPP 133
QY 128 VLVYVFGMEFRTNQPFKAMQGVQKIVNMKSENLFESQGGPIIQAQIENIEYGPVWE 187
Db 134 VLVYVFGVAFRTDNEPFAKMQKFTAKIVSMKAEKLFESQGGPIIQAQIENIEYGPV 193
QY 188 EIGAPGRKATKWAQMAVGLKTVGVPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPK 247
Db 194 EIGAPGRKATKWAQMAVGLKTVGVPWIMCKQEDADPPVIDTCNGFCYCEFRPNKPK 253
QY 248 WTEVMTGWTKFGGPIIPQPAEDIAFSAVAREVQNGSFYNYMHGCTNFGTSGLFIA 307
Db 254 WTEVMTGWTKFGGPIIPQPAEDIAFSAVAREVQNGSFYNYMHGCTNFGTSGLFIA 313
QY 308 TSVDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEAHV 367
Db 314 TSVDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLSGNSQEAHV 373
QY 368 GACAAFLSNYSRYSVKVTFQRPYNLPWISILPDKCTAVNTAQNVS-QSSSIKMT 426
Db 374 GACAAFLSNYSRYSVKVTFQRPYNLPWISILPDKCTAVNTAQNVS-QSSSIKMT 433
QY 427 AGGSLWSQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLVMTNVTNATSEGLFK 486
Db 434 VSSAFDMQSYNEETPADDDSTLTANGLEQKNVTRDSSDYLVMTNVTNATSEGLFK 493
QY 487 KDPVLTVMASAGHVLHVFVNGKLSGTYGTLDNPKLTVSGNVKLAGINKISLSVSVGL 546
Db 494 QYPLVTVMASAGHVLHVFVNGKLSGTYGTLDNPKLTVSGNVKLAGINKISLSVSVGL 553
QY 547 NVGVHDTWAGVLGPTVTLGSLNEGSRLAKQKWSYKVLGKESLSLHSLSGSSVVEW 606
Db 554 NVGVHDTWAGVLGPTVTLGSLNEGSRLAKQKWSYKVLGKESLSLHSLSGSSVVEW 613
QY 607 GSLMAQKQPLTWKATFNAPGNDPLALDMASMGKGQIWTNGEGVGRHWPYIAQGC 666
Db 614 GSSLVKKQPLTWKATFNAPGNDPLALDMASMGKGQIWTNGEGVGRHWPYIAQGC 673
QY 667 CSYAGTFNEKKQTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 723
Db 674 CNYAGTFNEKKQTCNQCPQSRQVHYPRSWLKPGLNLLVVEWGGNPTGISLVRRS 730

RESULT 8
ID 082670 PRELIMINARY; PRT; 723 AA.
AC 082670;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
OX NCBI_TaxID=3827;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED EPICOTYLS;
RA Dopic B., Esteban R., Labrador E.;
RT "the fourth beta-galactosidase clone found in Cicer epicotyls.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AJ011010; CA009457.1;
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Glycosidase; Hydrolase.
SQ SEQUENCE 723 AA; 80655 MW; 0585760D631B293A CRC64;

Query Match 75.1%; Score 2948.5; DB 10; Length 723;
Best Local Similarity 72.6%; Pred. No. 8.1e-191;
Matches 522; Conservative 88; Mismatches 106; Indels 3; Gaps 2;

QY 5 NVL-LLLVICLLDFSSVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLIQA 63
DB 6 NLGLMLLLCF--WCAVTASVTDHKTIVIDGQRRLISGSIHYPRSTPEWPAFLQKA 63

QY 64 KDGGLDVIETVYVWNGHEPSPGKYNFEGRYDLVRFIKWQVQAGLYVNLRIQGYVCAEWNF 123
DB 64 KEGGLDVIETVYVWNGHEPSPGKYNFEGRYDLVRFIKWQVQAGLYVNLRIQGYVCAEWNF 123

QY 124 GGFVPLWKYVPGMEFRINQPFVKAMQGFVOKIYNNMKSENLFESQGGPIIQAQIENYG 183
DB 124 GGFVPLWKYVPGISFRITDNEPFKAAMQKFTTKIYNNMKSENLFESQGGPIIQAQIENYG 183

QY 184 PVEVEIGAPGKAYTKAAQMAVGLTKGVPTMCKOEADPDVIDTCNGFYCEGFRPNKY 243
DB 184 PVEVEIGAPGKAYTKAAQMAVGLTKGVPTMCKOEADPDVIDTCNGFYCEGFRPNKY 243

QY 244 KPKMWTETVGTGWYTKFGGPIQRPADIAFVSARFVQNGNSFFNYMYHGGTNGFRSSG 303
DB 244 KPKMWTENWGWYTDGNAICVRYVEDLAYSVARFIQNRGFSVNYMYHGGTNGFRSSG 303

QY 304 LFIATSYDYDAPDEYGLLNEPKYCHLRLDHLKAIKLEPALVSSYAAVTSLSGNSQEAHVY 363
DB 304 LFIATSYDYDAPDEYGLTNEPKYCHLRLDHLKAIKLEPALVSSYAAVTSLSGNSQEAHVY 363

QY 364 RSKSGACAALSNYSYRVSVKVTQNPYPNLPWPSISILPDKCTAVYNTAQNVSQSSIK 423
DB 364 STGTSVCAAFLANVDTKSAATVTEGNGKYDLPWMSVILPDKCTDVNTAKVGAQSSQKT 423

QY 424 MTPAGGSLWSQSYNEETPTADSDTLTANGLEWQKNVTRDSSDYLVWNTVNYNIASNEGFL 483
DB 424 MISTNSTFDWQSYTEEPAFSEDDSDITAEALWEQINVTROSSDYLVWNTVNYNIASNEGFL 483

QY 484 KNGKDPVLYTNMAGHVLHVFNGLSGTVYGTLDNPKLTYSGNVKLRAGINKILLSVSV 543
DB 484 KNGYPIILNVMAGHVLHVFNGLSGTVYGTLDNPKLTFNSNVNLVAGNKNILLSVAV 543

QY 544 GLPNVGLHFEWNVGVLPVTLKGLNEGRDLSWQKNSYKVLKGESLSLHTITGGSSVD 603
DB 544 GLPNVGLHFEWNVGVLPVTLKGLNEGRDLSWQKNSYKVLKGESLSLHTITGGSSVD 603

QY 604 WVRGSLMAQKOPLTWYKATFNAPGNPDPLADMSMGKQIWIINGEGRHWPQYIAQGD 663
DB 604 WTQGSLLAKKOPLTWYKATFNAPAGNDPLGLDMSMGKGEIWNQDSTGRHWPQYIAHGS 663

QY 664 CSKCSYAGTNEKCOFNCGOPQSORWVHVPRLKPSGNLLVVEEWCNPTGISLYRR 722
DB 664 CGDCDYAGTFTNTKRCNTCGNPTQWYHIPKSWLNPTGNVLVLEEWGDPGSGISLLKR 722

RESULT 9
Q92P11 PRELIMINARY; PRT; 838 AA.
ID Q92P11
AC Q92P11;
```

```
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
GN TEGIA.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. MONEY MAKER;
RA De Silva J., Jarman C., Strongitharm B., Gidley M.;
RT "In vivo manipulation of plant cell walls.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AJ012796; CA010173.1;
DR InterPro: IPR000922; Gal_Lectin.
DR InterPro: IPR001944; Glyco_hydro_35.
DR Pfam: PF02140; Gal_Lectin; 1.
DR Pfam: PF01301; Glyco_hydro_35; 1.
DR PRINTS: PR00742; GLHYDRLASE35.
DR PRODOM: PD005612; Gal_Lectin; 1.
DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
DR PROSITE: PS02228; SUEL_LECTIN; 1.
KW Glycosidase; Hydrolase; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 838 SS-GALACTOSIDASE.
SQ SEQUENCE 838 AA; 92924 MW; AC2E11ABFA417762 CRC64;

Query Match 74.9%; Score 2941; DB 10; Length 838;
Best Local Similarity 72.0%; Pred. No. 3.2e-190;
Matches 520; Conservative 97; Mismatches 101; Indels 4; Gaps 2;

QY 1 MLRTNVLVLLVICLLDFSSVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLI 60
DB 6 ILMLNVLVLL---LGSVWFGSTASVYDHRRAIIVNGORRLISGSIHYPRSTPEWPGII 62

QY 61 QKAKDGLDVIETVYVWNGHEPSPGKYNFEGRYDLVRFIKWQVQAGLYVNLRIQGYVCAE 120
DB 61 QKAKEGGVDTVQTVVWNGHEPQOGKYFYEGRYDLVRFIKVHQAGLYVHLRVGPYCAE 122

QY 121 WNFEGFPVWLKYVPGMEFRINQPFVKAMQGFVOKIYNNMKSENLFESQGGPIIQAQIEN 180
DB 123 WNFEGFPVWLKYVPGISFRITDNGPFKAAMQKFTAKIYNNMKKAERLYETQGGPIILS 182

QY 181 EYGVVEVEIGAPGKAYTKAAQMAVGLTKGVPTMCKOEADPDVIDTCNGFYCEGFRPN 240
DB 183 EYGVMEVELGAPGKSYAQAAKMAVGLTGVPWVWCKQDDAPDPIINACNGFYCDYFSPN 242

QY 241 KPYKPKMWTETVGTGWYTKFGGPIQRPADIAFVSARFVQNGNSFFNYMYHGGTNGFR 300
DB 243 KAYPKIWTETAWTAFWTFGFGNPVYRPAEDLAFSAKFIQKGGSFINYMYHGGTNGFR 302

QY 301 SSGLFATSYDYDAPDEYGLLNEPKYCHLRLDHLKAIKLEPALVSSYAAVTSLSGNSQEA 360
DB 303 AGGFATSYDYDAPDEYGLLRQPKWGLKDLHRAIKLEPALVSGDPVAVTALHQOEA 362

QY 361 HVYRSKGACAALSNYSYRVSVKVTQNPYPNLPWPSISILPDKCTAVYNTAQNVSQSS 420
DB 363 HVFBSKAGSACAFLANDYDQHSFATVSEANRHYNLPWPSISILPDKCTVNTAIGAQA 422

QY 421 SIKMTAGGSLWSQSYNEETPTADSDTLTANGLEWQKNVTRDSSDYLVWNTVNYNIASNE 480
DB 423 QMKMTVPVSRGLPWSQSYNEETSSYEDS--SFTVVGLEQINTTRDVS DYLVWNTVNYNIASNE 481

QY 481 GFLKNGKDPVLYTNMAGHVLHVFNGLSGTVYGTLDNPKLTYSGNVKLRAGINKILLS 540
DB 482 KFLRGKWPMLTITMSAGHALHVFNGLQAGTAVGSLKPKLTFSAVNLNLRAGVNVKILLS 541
```


RC	STRAIN=CV. CHANDLER; TISSUE=FRUIT;
RA	Trainotti L., Spinello R., Casadoro G.;
RT	*Unusual beta-galactosidases with a lectin-like domain are expressed in strawberry.*
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBDJ databases.
DR	EMBL; AJ278703; CAC44500.1; -.
KW	Hydrolase; Glycosidase.
SQ	SEQUENCE 843 AA; 93505 MW; C02EFF2C93EDBA4A CRC64;
Query Match 74.3%; Score 2915; DB 10; Length 843;	
Best Local Similarity 71.8%; Pred. No. 1.8e-188;	
Matches 519; Conservative 95; Mismatches 101; Indels 8; Gaps	
Qy	5 NVLL--LLVLCILDFSSVKASVSVDRAIINGKRKILISGSIHYPRSTPMWPDLIQ 61
Dd	: : : : : : : : : :
Dd	9 NVVVAAALVLC--SCFASVRASVSYSKAIVNGQRILLISGSIHYPRSTPEMFDLIQ 66
Qy	62 KAKDGLDVITYVFVWNGHEPSPGKYNFEGRYDLVRFIKMVRAGLYVNLRIGRPYYCAEW 121
Dd	: : : : : : : : : : : :
Dd	67 RAKDGLDVITYVFVWNGHEPSPGKYFYFEDNYDLVKFIKLVOAGLVYVHLRIGRPYYCAEW 126
Qy	122 NFGCFPVWLKVYVGMCEPRTNNQPFKAMQGFQVKIYNMKSENLFPSQGGPIIMAOIENE 181
Dd	: : : : : : : : : : : :
Dd	127 NFGCFPVWLKVYVGIQTONGPFKDQMGRFTKIYNMKKAERLFFSHGQPILLSOIENE 186
Qy	182 YGPWEIEIGAPGKAYTKWAQAQAVGLKTGYPTWIMCKOEDAPDPVIDTCNCFYCCEGFRPNK 241
Dd	: : : : : : : : : :
Dd	187 YGPWEIEIGAPGKAYTDWAQAQAVGLCTGYPPWCMCKDDADPDVINACNGFYCDYFSPNK 246
Qy	242 PYKPKMTTEVMYTTFKEGGPIIPORPAEDIAFSAVARFVQNNGSFNNYMYHGGTNEFRGS 301
Dd	: : : : : : : : : : : :
Dd	247 AYKPKMTEAWTGWFTEFGGAVPYRPAEDLAFSAVKFLQKGGAFINYMYHGGTNEFRGT 306
Qy	302 SGLFIATSYDYDAPLDDEYGLLNPEPKYCHLDLHKATKLSEPALVSSYAATSLGSNQEAH 361
Dd	: : : : : : : : : : : :
Dd	307 GGPFIAATSYDYDAPLDDEYGLLRQPKWGHLKDLHRAIKLCEPALVSSDPTVTPLGTQEAH 366
Qy	362 VYRSKCACAFAFLSNYSRYSVKVTFEONRRPNLPWSISITLPCKTAIVNTAQVNSOSSS 421
Dd	: : : : : : : : : : : :
Dd	367 VFKSNSACAFAFLANYNRKSKFAVGNGMHYNLPWSISITLPCKTVTYNTARIGAQTAR 426
Qy	422 IKM--TPAGGSLMSQSYNETPTADDSDTLTANGLMEQKNVTRDSSDYLMWTNVNIASN 479
Dd	: : : : : : : : : :
Dd	427 MKMPRVPIHGFSQWAINDETATYSDTSETTA-GLLEQINITRDATDYLWYMTDKIDPS 485
Qy	480 EGFLKNGKDPYLTVMSAGHVHVFVNGKLSGTYGILDNPCLTYSGNVKLRAGINKISLL 539
Dd	: : : : : : : : : :
Dd	486 EDFURSGNYPVLTVLSAGHAIRVFINQLAGTAGYSLETPLKFPGQVNLRAGINOIAL 545
Qy	540 SVSVGLPNVGVHDVFNAGVLPVTTLSGLNEGSRNLAKOKWSYKVGILGESLSLHSISGS 599
Dd	: : : : : : : : : : : :
Dd	546 SIAGVLPNGVGHETWNAGILGPVILNGLNEGRRDSLQWKSWTKIGLKGEALSLSLTGS 605
Qy	600 SSVEWVRGSLMAQKPULTWYKATFNAPGGNDPLALDMA SMKQIQIWIINGEVGRHWPGVI 659
Dd	: : : : : : : : : :
Dd	606 SSVEWTEGSFVAQRQPLTWYKTTFNRPAGNSPLALDMGSMGQGVINDRSIGRYPAYK 665
Qy	660 AQGDSCSCSYAGTFNEKKOTNGCQPSQRVHYVPSRWLKPSPGNLLVVFEEWGGNPTGISL 719
Dd	: : : : : : : : : : : :
Dd	666 ASGTGCBCNTAGTFSEKKCLNSCGEASQRVHYVPSRWLNPTGNLLVYLEWGGDPNGIFL 725
Qy	720 VRR 722
Dd	
Dd	726 VRR 728
RESULT 12	
ID Q9Zp17	PRELIMINARY; PRT; 730 AA.
AC Q9Zp17	
DT 01-MAY-1999 (TrEMBLrel. 10, Created)	
DR 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)	

DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).
OS	Lupinus angustifolius (Narrow-leaved blue lupine).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae
OC	eurosid I; Fabales; Fabaceae; Papilionoideae; Genisteae; Lupinus.
OC	NCBI_TaxID=3871;
FX	[1]
RN	SEQUENCE FROM N.A.
RP	TISSUE=SEED ENDOSPERM;
RC	Chengappa S.;
RA	"Heterologous Expression of a galactan specific ss-galactosidase
RT	results in extensive changes in plant cell wall composition and
RT	rheology.";
RL	Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC	-!- CATALYTIC ACTIVITY: HYDROLYSTS OF TERMINAL, NON-REDUCING BETA-D-
CC	GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC	-!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR	EMBL; AJ011047; CAA09467.1; -.
DR	InterPro; IPR001944; Glyco_hydro_35.
DR	Pfam; PF01301; Glyco_hydro_35; 1.
DR	PRINTS; PR00742; GLHYDRASE35.
DR	PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW	Glycosidase; Hydrolase; Signal.
FT	SIGNAL 1 33
FT	CHAIN 34 730
FT	SEQUENCE 730 AA; 81686 MW; B69D3C64E9CD0E11 CRC64;
SQ	EXO GALACTANASE.

Query Match	74.18;	Score 2906;	DB 10;	Length 730;
-------------	--------	-------------	--------	-------------

Best Local Similarity	71.6%	Pred. NO. 6.1e-188;
Matches	521; Conservative	94; Mismatches 99; Indels 14; Gaps 5;

QV 1 MLRTN--VLLLV--ICLLDFFSSVKASVSYDDRAIINGKRKILISGSIHYPRSTPQM 55

Db 12 MSRRNFHVVLLLLFFWVCY-----VTASVTDHKAIMNGORRILISGSIHYPRSTPOM 65

56 WPDLIOAKDGGCLDVTETVVEWNGHEPSPGKYNEEGRYDI.VRETKMVORAGI.VVNI.RIGP 115

[illegible]

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

QY I/O AQ IENEIGF VENEIGAFORATIRNWEAQTA VGEKIGF VWINCQKQEDAF V I D I C N O F I C I 255

DB I86 S Q I E N E Y G P F V E W E I G A P G K A Y T R K W A A Q M A V G L D T G V P W V M C R A Q E D A P D P T I D T C N G F I C E 243

QY 236 GERPNKPYKPKMWTETVWTGWYTKFGGPIQRPAPEDIAFSVARFQNNGSFFENYMYHGGT 295

Db 246 NFTPKNYKPKLWTENWTGWYTAFGGATPYRPAEDIAFSVARFIQNRGSLFNYYMYHGGT 305

QY 296 NFGRTSSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSLG 355

Db 306 NEGRTSNGLEFVATSYDYDAPIDEYGLLNEPKWGLRELHRAIKQCESALVSVDPTVSWPG 365

OV 356 SNOEAHVYRSKSGACAAFLSNYDSRYSVKVTFQNRPNLPPWSISILPDCCKTAVYNTAQV 415

db 366 KNI.EVHI.VKTES-ACAAFLANYNTDYSTOVKEGNGOYDI.PPWSISILPDCKTEVENTAKV 424

416 NSOSESISKVTPDACCISWOSYNFEETPTADDSDTLTANGIWEOKNVTBDSSTYI.WYMTNVN 475

		:		:	:	:				: :		
		:		:	:	:				: :		
NORTH	WINDMILL	B	PICCOLI	COUNTDOWN	CONCERNED	CHUCK	F	STATION	NUMBER	CONFUSE	MURDER	484

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

[illegible]

QY 336 ISLTSVSGLFNVGVHYDITWNAAGVLGPVITSLGELNEGSRNLAKQWNSIKVGLRGESTLSHS 333

db 543 ISLLSVSVGLANVGTHTFEETWNTGVLGPVTLTGLSSGTWDL SKQKWSYK IGLRGESLSLHT 602

Qy	596	LSGSSSVVWVRGSLMAQKQPLTWYKATFNAPCGNDPLALDMSMGKGQIWINGEVGVRHW	655
Dd	603	EAGSNSEVWVGSLVAKKQPLAWYKTTFSPAGNDPLALDLGSMGKGEVYVWNGQSIGRW	662
Qy	656	PGYTAQDCSKCVAGTGFENKCKOTNCNGOPSQRWYHVPWSLKPSCNLLLVFEEWGNT	715
Dd	663	PGNKARGCNCNYAGTYDTCKLANCGQDSQRWYHVPWSLRSGGNLYLVLEEWGGDN	722
Qy	716	GISLVRRS 723	
Dd	723	GIALVERT 730	
RESULT 13			
Q93X56 PRELIMINARY; PRT; 722 AA.			
ID	AC	Q93X56;	
DT	01-DRC-2001 (T-EMBLrel. 19, Created)		
DT	01-DRC-2001 (T-EMBLrel. 19, Last sequence update)		
DT	01-DRC-2001 (T-EMBLrel. 19, Last annotation update)		
DE	BETA-GALACTOSIDASE (EC 3.2.1.23).		
GN	BETA-GAL3		
OS	Fragaria ananassa (Strawberry).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eucoids I; Rosales; Rosaceae; Rosoideae; Fragaria.		
OX	NCBIs_TaxID=3747;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-CV. CHANDLER; TISSUE=FRUIT;		
RA	"Ratnott L., Spinello R., Casadoro G.;		
RT	Unusual beta-galactosidases with a lectin-like domain are expressed		
RL	in strawberry."		
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases:		
DR	EMBL: AJ278705; CAC44502.1; .		
KW	Hydrolase; Glycosidase.		
SQ	SEQUENCE 722 aa; 81158 MW; DC08461E3F380123 CRC64;		
Query Match 73.9%; Score 2901.5; DB 10; Length 722;			
Best Local Similarity 71.6%; Pred. No. 1.2e-187;			
Matches 513; Conservative 100; Mismatches 100; Indels 3; Gaps			
Qy	8	LLAVLICLDFFSSVKASVSYDDRAIIINGKRKRILISGSIHYPSTPQMMPDLIQAKDGG	67
Dd	10	MFFLLFLVSLSSALASGVGDHRAILLVNGKRRIISGSIHYPSTPEMPPDLLQAKADGG	69
Qy	68	LQVIETTVFWNGHEPSGKYNFGRVDLVRFIKMVORAGLYVNLRTGPYVCAEWNFGGPP	127
Dd	70	LDVLQTIVFVWNGHEPSGKYFYEDRDYLKFIAQQHGLYHLIRGTPVCAEWNFGGPP	129
Qy	128	VWLKYVPCMEFRNNOPFKYAMOGVFQKIYNMKSENLFESQGPIIMAQIENEYGPVEW	187
Dd	130	VWLKYVPGIATFDNRDPFMAMEKEFTQKIYVMKABERLFTQGGPIILSOIENEYGPVEW	189
Qy	188	EIGAPGRAYTKWAQAQMAVGLKTGPWTMCKOEDAPDPVIDTCNGFCYCEGRPNPKYKPKM	247
Dd	190	EIGAPGKSYTQWAKMAVGLNTGVPWMCKOEDAPDIIDTCNGFCENFTPNKNYKPKM	249
Qy	248	WTEVWTGWYTKFGGPIIPQPAEDIAFSARFVNQNSFFNYMYHGHTNFGRTSSGLFTA	307
Dd	250	WTEIWTGWYTEFGAVPTREPAQLATSFVARFIQNGGSFANYMYHGHTNFGRTAGGPFIA	309
Qy	308	TSYDYDAPLDEYGLLNPEPKYCHLRDLHKALKSEPALVSSYAAYVTSLGNSQEAHVYRSK	367
Dd	310	TSYDYDAPLDEYGLLPFPKYSHLKYMHKAIKMAEPALLATDAAVSKLGNNQEAHVYQRS	369
Qy	368	GACAALFNSYDSRVSKVTQNRPYNLPWPMSISITLDPCKTAVYNTAQVNSQSSSIKWTPA	427
Dd	370	G-AAFLANLYDTKPVPRVTTWNKOYNLPPMSISITLDPCKTEVENTARY-QQSPPTKMTPV	427
Qy	428	GGGLSWQSYNEETPTADDSTLTANGLWEOKNVRTDSSDYLYWMTNVNIASNEGFLKNKG	487

Db	428	-AHL5WQAYIEDVAT5ADDAFTSVGLRQEI5U5TWDNTDYLWTWTDIT1GPNQEQFLRTQK	488
Qy	488	DPYLTVMSAGHLVHFVNGKLSGTVYGTLDNPXLTYSGNVKLRAGINKLSLSVSYGLEN	547
Db	487	YPTLKVDSAGHALHVFINGQLSGSAYGTIAFPKLEFNQGVKLRAGINKLALLSVSYGLAN	546
Qy	548	GVHYDTWNAAGVLGPVYTLISLNGESRNLAQKQWSYKVLKGESLSLHSLSGSSSVWVRG	607
Db	547	VGLHFEFTWNTGVLPVTLVAGVSGTWDMTKQWYTKIGMRGEDMSLHTVSGSSSVWVQG	606
Qy	608	SLMAQKPLTWYKATFNAPGNDPLALDMSMGKGOIWIINGEGVRHWPYTAQGDCSKC	667
Db	607	SLIAQYRPLTWYKAILNAPGNAFLALDMSMGKGOIWIINGEGVRHWPYTAQHSQCGAC	666
Qy	668	SYAGTFNEKKCQTCNGQPSQRWYHVPWSLWKP5GNLLVVFEEWGGNPTGISLVRRS	723
Db	667	YIACYYTENKRTNCGQPSQRWYHVPWSLWKP5GNLLVVFEEWGGDPTKISLVARS	722
RESULT 14			
Qy	Q9SCV0	PRELIMINARY; PRT; 728 AA.	
AC	Q9SCV0;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE).		
GN	BGAL12.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RA	Gy I., Kreis M., Lecharny A.;		
RT	"The Beta-galactosidases are encoding by a multigene family in		
RT	Arabidopsis thaliana."		
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.		
CC	-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-		
CC	GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.		
CC	EMBL: AJ270308; CAB64748.1; -		
DR	InterPro: IPR001944; Glyco_hydro_35.		
DR	Pfam: PF01301; Glyco_hydro_35; 1.		
DR	PRINTS; PR00742; GLHYDRLASE35.		
DR	PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.		
KW	Glycosidase; Hydrolase; Signal.		
FT	SIGNAL 1 23 POTENTIAL.		
FT	CHAIN 24 728 PUTATIVE BETA-GALACTOSIDASE.		
SQ	SEQUENCE 728 AA; 81532 MW; 75EB0999147BA65 CRC64;		
Query Match 72.6%; Score 2848.5; DB 10; Length 728;			
Best Local Similarity 69.8%; Pred. No. 4.6e-184;			
Matches 502; Conservative 104; Mismatches 108; Indels 5; Gaps			
Qy	7	LLLLVLCILDFSSVKASVYDRAIIINGKRKILISGSIHYPRSTQMPWPDLIQAKDG	66
Db	11	ILLGLICSSLLCSKVAIVYDRKAVLINGQRRIILSGSIHYPRSTQMPWPDLIQAKDG	70
Qy	67	GLDVIYTVFWNGHEPSPGKYNFEGRYDLYRFITKMYORAGLYNLRITGPYVCAEWNFGGF	126
Db	71	GLDVIQIYVFWNGHEPSPGQYFYEDRDYDLVKFIFKVQOAGLYVHLRIGPVCVCAEWNFGGF	130
Qy	127	PVWLKYPVGMFEFRNNOPFKVAMQGFQYKQITVNMNMKSENLFESQGGPIIMAIQENYGPVE	186
Db	131	PVWLKYPVGMVFRTDNEPEFAAMQKFEKIVRMKMEKIFETQGGPIILSQIENYGPTE	190
Qy	187	WEIGAPKATYKRAQMAVGLKTGVPMWCKQEDADPPVIDTNCNGFYCSGFRPNKYPKPK	246
Db	191	WEIGAPKATYKWAEMAQGLSTGVPMWCKQEDADPPVIDTNCNGFYCSGFRPNKYPKPK	250

```
QY 247 MWTEVWTHYTKFGGPIQORPAEDIAFSVARFVONNGSFENYMYHGGNFRTSSGLFI 306
Db 251 MWTEVWTHYTKFGGPIQORPAEDIAFSVARFVONNGSFENYMYHGGNFRTSSGLFI 309
QY 307 ATSYDYDAPLDEYGLNPEKYGHLRDLHKAIKLSEPALVSSAAVTSLSNQEAHYRSK 366
Db 310 ATSYDYDAPLDEYGLNPEKYGHLRDLHKAIKLSEPALVSSAAVTSLSNQEAHYRSK 369
QY 367 SGCAAFSLNYSRYSVKVTFQNRPNLPPWSISILPDCCKTAVYNTAQNSSSST--KM 424
Db 370 S-SCAAFLSNYNTSSAARVLFGGTYDLPWWSVILPDCCKTAVYNTAKVQVTSIHHMKM 428
QY 425 TPAGGLSWQSYNEETPTADSDTLTANGLWQKNTVRSDDLWYMTNVTNVIASNEGLK 484
Db 429 VPTNTPFSWGSYNEEIPSSANDNGTFSQDGLVQISITRDKTDYFWLTDITISPEKFL- 487
QY 485 NGKDPYLTVMASAGHLVHFVNGKLSGTVYGLDNPKLTYSGNVKLRAGINKISLSVSG 544
Db 488 TGEDPLLTIGSAGHALVHFVNGKLSGTVYGLDNPKLTYSGNVKLRAGINKISLSVSG 547
QY 545 LPNVGVHYDTWAGVLPVTLTSLNGESRNLAKQKWSYKVGKLSLHSLSGSSSVV 604
Db 548 LPNVGVHYETWNTGVLGPTLVNGVNSGTWDMTKKWSYKIGTKGEALSHTLAGSSTVEM 607
QY 605 VRGSLMAQKQPLTWKATFNAGGNDPLALDMASMGKQIWIINGEGVGRHWPYIAQGD 664
Db 608 WKEGSLVAKKQPLTWKATFNAGGNDPLALDMASMGKQIWIINGEGVGRHWPYIAQGD 667
QY 665 SKCSYAGTENEKKCOTNCGQPSQWYHVPWSLKPNSGNTLVVFEWGGNPTGISLVRS 723
Db 668 ERCSTAGTTEKKCLSNCGEASQWYHVPWSLKPNSGNTLVVFEWGGNPTGISLVRS 726

RESULT 15
Q9SZ15 PRELIMINARY; PRT; 729 AA.
AC Q9SZ15;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN F20818.250 OR AT4G26140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Rose M., Hempel S., Entian K.-D., Hoheisel J., Mewes H.W.,
RA Mayer K.F.X., Schueller C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
DR EMBL; AL049483; CAB39679.1; -
DR EMBL; AL161564; CAB79469.1; -
DR InterPro; IPR001944; Glyco_hydro.35.
DR Pfam; PF01301; Glyco_hydro.35; 1.
DR PRINTS; PR00742; GLHYDRLASE3.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
```

KW Glycosidase; Hydrolase.
SQ SEQUENCE 729 AA; 81660 MW; F5B75526720F696D CRC64;

Query Match 72.3%; Score 2838; DB 10; Length 729;
Best Local Similarity 69.7%; Pred. No. 2.4e-183;
Matches 502; Conservative 104; Mismatches 108; Indels 6; Gaps 5;

```
QY 7 LLLLVICLLDFFSSYKASVSYDDRAILLNGKRLKILSGSIHYPRSTPOMWPDLOKAKDG 66
Db 11 ILGLICCCSLSCSKAIVITDRKAVIINGQRILLSGSIHYPRSTPOMWPDLOKAKDG 70
QY 67 GLDVIETVYFWNGHEPSPGKYNFEGYDLVRFIKVQVAGLVNLRIGRYPVCAENFNGGF 126
Db 71 GLDVIQTVVFWNGHEPSPGQYVYFEDRYDLVKFKVQVAGLVNLRIGRYPVCAENFNGGF 130
QY 127 PWLVKIVVGMERTNNOQPKVAMQGFVQKIVNMKSENLFESQGGPIIMAOIENEGPVE 186
Db 131 PWLVKIVVGMERTNNOQPKVAMQGFVQKIVNMKSENLFESQGGPIIMAOIENEGPVE 190
QY 187 WEIGAPGKAYTKWAAMAVGLKTGVPWIMCKOEDAPDPVIDTCNGFYCEGPRPNKPYKPK 246
Db 191 WEIGAPGKAYTKWAAMAVGLKTGVPWIMCKOEDAPDPVIDTCNGFYCEGPRPNKPYKPK 250
QY 247 MWTEVWTHYTKFGGPIQORPAEDIAFSVARFVONNGSFENYMYHGGNFRTSSGLFI 306
Db 251 MWTEVWTHYTKFGGPIQORPAEDIAFSVARFVONNGSFENYMYHGGNFRTSSGLFI 309
QY 307 ATSYDYDAPLDEYGLNPEKYGHLRDLHKAIKLSEPALVSSAAVTSLSNQEAHYRSK 366
Db 310 ATSYDYDAPLDEYGLNPEKYGHLRDLHKAIKLSEPALVSSAAVTSLSNQEAHYRSK 369
QY 367 SGCAAFSLNYSRYSVKVTFQNRPNLPPWSISILPDCCKTAVYNTAQNSSSST--KM 424
Db 370 S-SCAAFLSNYNTSSAARVLFGGTYDLPWWSVILPDCCKTAVYNTAKVQVTSIHHMKM 428
QY 425 TPAGGLSWQSYNEETPTADSDTLTANGLWQKNTVRSDDLWYMTNVTNVIASNEGLK 484
Db 429 VPTNTPFSWGSYNEEIPSSANDNGTFSQDGLVQISITRDKTDYFWLTDITISPEKFL- 487
QY 485 NGKDPYLTVMASAGHLVHFVNGKLSGTVYGLDNPKLTYSGNVKLRAGINKISLSVSG 544
Db 488 TGEDPLLTIGSAGHALVHFVNGKLSGTVYGLDNPKLTYSGNVKLRAGINKISLSVSG 547
QY 545 LPNVGVHYDTWAGVLPVTLTSLNGESRNLAKQKWSYKVGKLSLHSLSGSSSVV 603
Db 548 LPNVGVHYETWNTGVLGPTLVNGVNSGTWDMTKKWSYKIGTKGEALSHTLAGSSTVEM 607
QY 604 WVRGSLMAQKQPLTWKATFNAGGNDPLALDMASMGKQIWIINGEGVGRHWPYIAQGD 663
Db 608 WKEGSLVAKKQPLTWKATFNAGGNDPLALDMASMGKQIWIINGEGVGRHWPYIAQGD 667
QY 664 CSKCSYAGTENEKKCOTNCGQPSQWYHVPWSLKPNSGNTLVVFEWGGNPTGISLVRS 723
Db 668 ERCSTAGTTEKKCLSNCGEASQWYHVPWSLKPNSGNTLVVFEWGGNPTGISLVRS 727
```

Search completed: June 24, 2002, 20:42:04
Job time: 264 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:28:45 ; Search time 3524.51 Seconds
(without alignments)
9780.430 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	754.6	29.5	793	10	BI422837
2	754.2	29.5	778	10	BI423180
3	747.8	29.3	762	10	BM413054
4	665.6	26.1	704	10	BI432727
5	645.8	25.3	649	10	BF113419
6	645.6	25.3	653	9	AW030222
7	618.8	24.2	622	9	AW096565
8	613.6	24.0	620	10	BM410481
9	586.4	23.0	596	9	AW223645
10	581	22.7	589	9	AW029855
11	556.6	21.8	563	9	AW032987
12	549.4	21.5	551	10	BE431542
13	537.8	21.1	549	9	AW033711
14	534.4	20.9	536	9	AW035517
15	532.4	20.8	611	9	AI894883
16	524.4	20.5	526	10	BE436849
17	512.8	20.1	533	9	AW616819

18	503.8	19.7	528	9	AW033293	AW033293 EST276864
19	497	19.5	497	9	AW029823	AW029823 EST273078
20	490.4	19.2	494	9	AW217169	AW217169 EST295883
21	474.8	18.6	478	10	BE434106	BE434106 EST405184
22	473.8	18.6	612	9	AW030607	AW030607 EST273862
23	472.8	18.5	612	9	AW030428	AW030428 EST273683
24	452.4	17.7	455	9	AW441477	AW441477 EST310873
25	450.2	17.6	612	10	BF187952	BF187952 EST444239
26	435.6	17.1	611	10	BG097633	BG097633 EST462152
27	423.4	16.6	661	10	BF641354	BF641354 NF066611
28	421	16.5	647	10	BI922495	BI922495 EST542399
29	414.6	16.2	610	9	AI894508	AI894508 EST263951
30	406	15.9	481	9	AW617501	AW617501 EST323912
31	404	15.8	738	10	BG841036	BG841036 MEST14-H0
32	403.8	15.8	659	9	AW696211	AW696211 NF103F08S
33	398.6	15.6	405	9	AI894534	AI894534 EST263977
34	393	15.4	789	10	BG123351	BG123351 EST468997
35	378.8	14.8	728	10	BI310794	BI310794 EST531254
36	377.8	14.8	769	10	BM412213	BM412213 EST586540
37	373.4	14.6	535	9	AB075003	AB075003 AB075003
38	370.4	14.5	674	9	AI778902	AI778902 EST259781
39	367.8	14.4	379	9	AW029933	AW029933 EST273188
40	365.4	14.3	726	10	BE660167	BE660167 O-D10 Gna
41	365.4	14.3	800	10	BG645654	BG645654 EST507273
42	364.4	14.3	604	10	BF632221	BF632221 NF029H10D
43	360	14.1	766	10	BI176439	BI176439 EST517384
44	358.4	14.0	666	9	AI055468	AI055468 coau0004B
45	358	14.0	358	9	AW933150	AW933150 EST358993

ALIGNMENTS

RESULT	1
BI422837	
LOCUS	BI422837
DEFINITION	EST533503 tomato callus, TAMU Lycopersicon esculentum cDNA clone
ACCESSION	BI422837
VERSION	BI422837.1 GI:15197578
KEYWORDS	EST.
SOURCE	tomato.
ORGANISM	Lycopersicon esculentum
REFERENCE	1 (bases 1 to 793)
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Lilang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE	Generation of ESTs from tomato callus tissue
JOURNAL	Unpublished (1999)
COMMENT	Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html .
FEATURES	source location/Qualifiers 1..793 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLEC71F18" /clone_lib="tomato callus, TAMU" /tissue_type="callus" /dev_stage="25-40 days old" /lab_host="XL1-Blue MRF" /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included

BASE COUNT	238 a	159 c	177 g	219 t	undifferentiated masses. Tomato Callus EST Library"				
ORIGIN									
Query Match	29.5%;				Score 754.6;	DB 10;	Length 793;		
Best Local Similarity	98.1%;				Pred. No. 6.7e-96;				
Matches 774;	Conservative	0;	Mismatches 14;	Indels 1;	Gaps 1;				
QY	1051	ttgagagacttacataaagctatcaagctatctgaacggtgttagtttcatcatgct	1110						
Db	5	TTGAGAGACTTACATAAAGCTATCAAGCTATCTGAACCGGCTTTAGTTTCATCATATGCT	64						
QY	1111	gcggtgactagtcttgaagtcaatcaagagctcatgttttatagatacaaaatctggagct	1170						
Db	65	CGGTGACTAGCTCTTGAAGTAATCAAGAGCTCATGTTTATAGATCAAAATCTGGAGCT	124						
QY	1171	tgctgctcttttttataccaactatgactctagatattcagtaaaagtcacctttcagaat	1230						
Db	125	TGTGCTGCTTTTATCCAACTATGACTCTAGATATTTCAGTAAAGTCACTTTTCAGAA	184						
QY	1231	agccatacaaatctgctccatcgaatgcacatgatttcccgaactgcataaactgcgctt	1290						
Db	185	AGCCATACAACTGCTCCTCCTCATGCTCAGCATCTTCCCGACTGCAAAACTGCCGTT	244						
QY	1291	tacaacactgcacaggtttaactctcaagctcgagcataaagatgagcctcgaggtggt	1350						
Db	245	TACAACACTGCACAGGTAACTCTCAAGCTCGAGCATAAAGATGACGCTTCAGAGTGGT	304						
QY	1351	ggattctcttgccagtcatacaatgaagaaacgcctactgctgtagcagcgatacactt	1410						
Db	305	GGATTCTGTTGGCAGTCATACAATGAAGAAACGCTACTGCTGATGACAGCGATACACT	364						
QY	1411	acagctaaacgactatggaaacagaaaacgtcacagagattcatcagactatctgtg	1470						
Db	365	ACAGCTAAACGGACTATGGGAACAGAAAACGTCACAAGAGATTTCATCAGACTATCTGTGG	424						
QY	1471	tacatgacaaatgaaatagcatctaatgaaggatttctaaagaacggaaagatcct	1530						
Db	425	TACATGACAAATGTAAATATAGCATCTAATGAAGGATTTCTAAAGAACGGAAAGGATCCT	484						
QY	1531	tatctcaactgttatgcgctggtcatgtcttgcagtgcttcttcgaatggaaaaactatca	1590						
Db	485	TATCTCACTGTATGTCGCTGGTCATGCTCTTGATGTTTTCGTCATGGAAAACTATCA	544						
QY	1591	ggactgtttatggtacattggataatccaaaacttacatacagtggtgcaacggtgaagtta	1650						
Db	545	GGAACGCTTTATGGTACATTTGGATAATCCAAAACCTTACATACAGTGGCAACGTGAAGTTA	604						
QY	1651	agagctggattaaacaagatttctctcagtgcttccggttggtctcccg-aacgttgg	1709						
Db	605	AGAGCTGGTATTAAACAAGATTTCTCTGCTCAGTGTTTCCGTTGGTCTCCGAAAAACGTGG	664						
QY	1710	cgtgcattatgatacatggaatgcagaggttctaggtccagtcacgttgagcggctctcaa	1769						
Db	665	CGTGCATTATGATACATGGAATGCAGAGTCTTAGGGTCAGTCACTGTTGAGCGGGTCTCA	724						
QY	1770	tgaaggttcaagaacttggcgaacacgaaatggtcttcaaggttggctgaaagcgga	1829						
Db	725	TGAAGGGGTCAGAAACTTGGCAACAGAAATGGTCTTTACAAGTTGGTCTGTGAAGAGCGGA	784						
QY	1830	atcggttaag	1838						
Db	785	ATCGTTAAG	793						
RESULT	2								
LOCUS	BI423180								
DEFINITION	EST533846 tomato callus, TAMU Lycopopersicon esculentum cDNA clone								
ACCESSION	BI423180								
VERSION	BI423180.1 GI:15198218								

KEYWORDS	EST.														
SOURCE	tomato.														
ORGANISM	Lycopersicon esculentum	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.													
REFERENCE	1 (bases 1 to 778)														
AUTHORS	Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.														
TITLE	Generation of ESTs from tomato callus tissue														
JOURNAL	Unpublished (1999)														
COMMENT	Contact: CUGI	Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . Location/Qualifiers 1. .778 /organism="Lycopersicon esculentum" /cultivar="TA496" /db_xref="taxon:4081" /clone="cLEC76119" /clone_lib="tomato callus, TAMU" /issue_type="callus" /dev_stage="25-40 days old" /lab_host="Xl1-Blue MRP" /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"													
BASE COUNT	242 a 144 c 199 g 193 t														
ORIGIN															
Query Match	29.5%; Score 754.2; DB 10; Length 778;														
Best Local Similarity	98.8%; Pred. No. 7.7e-96;														
Matches 770; Conservative	0; Mismatches 8; Indels 1; Gaps 1;														
QY	1587									atcaggaactgtttatggtacattgataatccaaactacatacagtggaacgtgaa	1646				
Db	1									ATCAGGAACGTGTTATGGTACATTTGGATAATCCAAAACCTTACATACAGTGGCAACGTGA	60				
QY	1647									gttaagagctgtattacaagaattctctgctcaggtttcgttctcccgaa	1706				
Db	61									GTTAAGAGCTGGTATTTAACAAGATTTCTCTGCTCAGTGTTTCCGTTGGTCTCCCGAACGT	120				
QY	1707									tggcgtgcattatgatacatggaatgcaggagttcttaggtccagtcacgttgagcggctt	1766				
Db	121									TGGCGTGCATTATGATACATGGAAATGCAGAGTCTTAGTCCAGTCACTGTTGAGCGGTCT	180				
QY	1767									caatgaaggttcaagaacttggcgaacacagaaatggtttacaagttggttgaagaag	1826				
Db	181									CAATGAAGGTCGAAGAACTTGGCGAAACAGAAATGGTCTTACAAGGTTGGTCTGAAAGG	240				
QY	1827									cgaatcgttaagtcttcactccttaagtggagttctctgttgaatgggttcagagttc	1886				
Db	241	CGAATCGTTAAGTCTTCACTCTCTTAAGTGGAGTCTTCTGTGTTGAATGGTTCGAGGTTT	300												
QY	1887	actaatggtcctcaaaacgagccctgacttggtacaaggtctacatttaacgcgcctggag	1946												
Db	301	ACTAGTGGCTCAAAACAGACGCCCTGACTTGGTACAGAGGTACATTTAAACGCCGCTGGAG	360												
QY	1947	aaatgatccactagcttttagacatggcaagtatgggaaaggtccaatattggataaatgg	2006												
Db	361	AAATGATCCACTAGCTTTAGACATGCCAAGTATGGGAAAAGGTACAGATATGGATAAATGG	420												
QY	2007	tgaagcgtagctcgccattgacctggatacatagcaagaagcgactgcagcaaatgcag	2066												
Db	421	TGAAGCGTAGTTCGCGCAATGGCCTTGGATACATAGCAACAAGCGCACTGCAGCAAAATGCAG	480												

[illegible]

Db	541	ACAGAGATGGCTCTACAAAGGTTCGCTAAAGGCGAATCGTTAAAGTCTTCACTCCTTAAG	600
QY	1854	tgggagttcttctgttgaaatgggttcgaggttcaactaatggctcaaaagcagcccttgac	1913
Db	601	TGGGAGTTCTTCTGTTGAATGGATTCGAGGTTCAGGTTCAGTTCGCTCAAAAGCAGCCTCTGAC	660
QY	1914	ttggtacaaggtacattttaacgcgcctggaggaaatgatccac	1957
Db	661	TTGGTCAAGGCTACATTTTAAACGCGCTGGAGGAATGAACCCAC	704
RESULT	5		
LOCUS	BF113419		
DEFINITION	EST441009 tomato breaker fruit Lycopersicon esculentum cDNA clone	649 bp	linear
ACCESSION	BF113419		EST 18-MAY-2001
VERSION	BF113419.1		
KEYWORDS	EST		
SOURCE	tomato.		
ORGANISM	Lycopersicon esculentum		
REFERENCE	1 (bases 1 to 649)		
AUTHORS	Alcala, J., Vrebalov, J., White, R., van der Hoeven, R.S., Holt, I.E.,		
	Liang, F., Hansen, T.S., Craven, M.B., Bowman, C.L., Bowring, C.M.,		
	Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley		
	, S.D.		
TITLE	Generation of ESTs from tomato fruit tissue, breaker stage		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: CUGI		
	Clemson University Genomics Institute		
	Clemson University		
	100 Jordan Hall, Clemson, SC 29634, USA		
	Email: http://www.genome.clemson.edu/orders/index.html .		
FEATURES	Location/Qualifiers		
source	1..649		
	/organism="Lycopersicon esculentum"		
	/cultivar="TA96"		
	/db_xref="taxon:4081"		
	/clone_lib="CLEG44D6"		
	/clone_type="tomato breaker fruit"		
	/tissue_type="pericarp"		
	/dev_stage="breaker"		
	/lab_host="SOLR"		
	/note="Vector: pBluescriptSkmCuadapt; Site_1: EcoRI;		
	Site_2: XhoI; supplier: Boyce Thompson Institute;		
	sequencing: The Institute for Genomic Research. Fruit		
	were harvested at the breaker stage (first sign of		
	lycopene accumulation on the blossom end of fruit). Fruit		
	were cut in half and the seeds and locules were discarded		
	prior to freezing the pericarp."		
BASE COUNT	198 a 110 c 151 g 190 t		
ORIGIN			
Query Match	25.3%;	Score 645.8;	DB 10; Length 649;
Best Local Similarity	99.7%;	Pred. No. 9.7e-81;	
Matches 647; Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
QY	291	ctggaatggacatgagccttctcctggaaaaataataatttgaaggagaatgatcttgt	350
Db	1	CTGGAATGGACATGAGCCTTCTCTCGAAATAATAATTTTCAGGAAGATGATGATCTGT	60
QY	351	tagattcatcaaatgatacagaagacagactttatgccaattacgtattggccctta	410
Db	61	TAGATTCAATCAAAATGTCACAGAGCAGGACCTTATGTCAATTTAGTATGTCCTTA	120
QY	411	cgtctgtgctgaatggaactttgggggattccctgtttggctaaaaatatgctgggtat	470
Db	121	CGPCTGTCGTGAATGAACATTTGGGGATTCCTGTTGGCTAAATATGATGTCCTGGTAT	180

QY 471 ggaattagaacaaacacgctttaaagtggtgctatgcaagattgttcagaaat 530
 |||||
 Db 181 GGAATTTAGAACAAACATCACGCTTTTAAAGTGGCTATGCAAGGATTGTTCACAAAT 240
 |||||
 QY 531 agtcaacatgatgaagtcagaaaaattgtttgaatcctcaagagagacaaataattatggc 590
 |||||
 Db 241 AGTCAACATGATGAAGTCAGAAAATTTGTTGAATCTCAAGGAGGACCAATAATTATGCG 300
 |||||
 QY 591 ccagatacaaaatgatgtagcaccagtagaattggaaattgctcctcggttaaagctta 650
 |||||
 Db 301 CCAGATAGAAAATGATGATGGACCAAGTAGAATGGAAATGGTCTCCCTGGTAAAGCTTA 360
 |||||
 QY 651 tacaataaggcgagctcaaaatggctgtagggtttgaaaactgggtgcccatgcatatg 710
 |||||
 Db 361 TACAATAATGGGAGCTCAAAATGGCTGTAGTGTGAAACTGGTGTCCCATGGATCATG 420
 |||||
 QY 711 taagcaagagatgctcctgactgctgtgattgatactgtaattgcttactatgcaagg 770
 |||||
 Db 421 TAAGCAAGAGGATGCTCTGATGCTGATGCTGATGATGCTGATGCTGATGCTGATGCT 480
 |||||
 QY 771 gtccgtccttaagccttcaaacctcaaaatgagacagcagcagcagcagcagcagcagc 830
 |||||
 Db 481 GTTCGCTCTTAATAGCCCTTCAAAACCTTAAATGTTGACAGAGGATGGACTGGCTG 540
 |||||
 QY 831 tacgaattcgtgtgtccaaatcctcaagaccagcagcagcagcagcagcagcagcagcagc 890
 |||||
 Db 541 TACGAATAATCGGTGTGCTCAAAATTCCTCAAAAGACCGACCGAAGACATGCAATTT 600
 |||||
 QY 891 caggtgtgtcagaacaaatggtcattcttctcaattactatcatatcatatcatatcat 939
 |||||
 Db 601 CAGGTTTGTTCAGAACAAATGGTTCAATCTCTCAATTTCAATTACTATCATATCAT 649
 |||||

RESULT 6

AW030222 653 bp mRNA linear EST 18-MAY-2001
 LOCUS EST273477 tomato callus, TAMU Lycopersicon esculentum cDNA clone
 DEFINITION cLEC19L18, mRNA sequence.

ACCESSION AW030222

VERSION AW030222.1 GI:5888978

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

REFERENCE 1 (bases 1 to 653)
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.

TITLE Generation of ESTs from tomato callus tissue
 JOURNAL Unpublished (1999)
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>

FEATURES 5 prime sequence.

Location/Qualifiers

1..653

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone="cLEC19L18"

/clone_lib="tomato callus, TAMU"

/tissue_type="callus"

/dev_stages="25-40 days old"

/lab_host="XLI-Blue MRF"

/note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:

XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons

of seedlings 7-10 days post-germination were excised, cut

at both ends and placed on MS medium with no selection.
 Mixed callus was harvested at 25 and 40 days and included
 undifferentiated masses. Tomato Callus EST Library"

BASE COUNT 186 a 123 c 174 g 169 t 1 others
 ORIGIN

Query Match 25.3%; Score 645.6; DB 9; Length 653;
 Best Local Similarity 99.2%; Pred. No. le-80;
 Matches 648; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1553 gtcagtcttgcattgtttctgcataatggaactatcaggaaactgtttatgtacattgg 1612

Db 1 GTCATGCTCTGTGATGTTTTCGTCATATGGAATACTATCAGGAACCTGTTATGTGTACATTGG 60

QY 1613 ataatccaaacttcacacagtcagcaactgaactgaagtaagactgtattatacaagattt 1672

Db 61 ATAAATCCAAACATTCATACAGTGGCAAGCTGAAGTTAAGAGCTGTTATTAACAAGATT 120

QY 1673 ctctgctcagtgcttccgttggtctcccgaacgttgctgcattatgatacatgaaatg 1732

Db 121 CTCTGCTCAGTGTTCCTGTTGGTCTCCCGAACGTTGGCGTCATTATGATACATGGAATG 180

QY 1733 caggagttctagttcagtcacgttgagcgtctcctcaatgaagggtcaagaacttgccga 1792

Db 181 CAGGAGTTCTAGTGTCCAGTCAGTTCAGCGGTCTCAATGAAGGGTCAAGAACTTGGCGA 240

QY 1793 aacagaataatggtcttcacaaggttgctcgaagggaactcgttaagcttcactccttaa 1852

Db 241 AACAGAAATGGTCTTTACAAGGTGGTCTGAAAGGGCAATCGTTAAGTCTTCACTCCTTAA 300

QY 1853 gtggaggtctctctgttggaatgggttcgaggttcactaatggtcctcaaaagcagccctga 1912

Db 301 GTGGGAGTTCTTCTGTTGAATGGTTCGAGGTTCACTAGTGGCTCAAAAGCAGCCCTGA 360

QY 1913 ctgtgtacaaggctacatttaacgcgcctggaggaaatgatccactagtttagacatgg 1972

Db 361 CTGTGTACAAAGCTACATTTAAACGCGCCTGGAGGAAATGATCCACTTAGCTTTAGACATGG 420

QY 1973 caagtatggaaaagtcagatatgataaattggtgaagcagtagtcgccattggcctg 2032

Db 421 CAAGTATGGGAAAGTCAGATATGGATATGATGTAATGGTGAAGCGGTAGGTCCCATTTGCCCTG 480

QY 2033 gatacatagcaaacgagcagctcagcaaatgagttatgctggaacgttcaacgagaaga 2092

Db 481 GATACATACACAAAGCGAGCTGCAGCAATATGAGTTATGCTGGAACTTCAACAGAGAAGA 540

QY 2093 agtcggaactaactcgcgcagcaacctctcagagatggttaccatgttccacgacatcg 2152

Db 541 AGTCCAGAGCTAACTCGCGACAACTTCTCAGAGATGGTACCATTGTTCACGATCGTGGC 600

QY 2153 tgaacacaaagtgaactgttagtatttcgaagaatggagagtaatcca 2205

Db 601 TGAACCAAGTGGAAACTTGTGTAGTAGTATTCGAAGATGGGAGGTAATCCA 653

RESULT 7

AW096565

LOCUS AW096565

DEFINITION

clone cLET38P6, mRNA sequence.

ACCESSION

AW096565

VERSION

AW096565.1

GI:6062160

KEYWORDS

EST.

SOURCE

tomato.

ORGANISM

Lycopersicon esculentum

Eukaryota; Viridiplantae;

Spermatophyta; Magnoliophyta;

Asteridae; euasterids I;

Lycopersicon.

REFERENCE

1 (bases 1 to 622)

AUTHORS

D'Ascenzo,M., He,X., Lyman,J., Holt,I.E., Liang,F., Upton,J.,
 Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,C.L., Nierman,W.,

TITLE
JOURNAL
COMMENT
Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,S.D. and Giovannoni
J.
Generation of ESTs from tomato leaf tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source

Location/Qualifiers
1. .622
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone_lib="cLEN38P6"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site 1: EcoRI; Site 2:
XhoI; cLET - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."
204 a 79 c 129 g 210 t

BASE COUNT
ORIGIN

Query Match 24.2%; Score 618.8; DB 9; Length 622;
Best Local Similarity 99.7%; Pred. No. 5.6e-77;
Matches 620; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 8 gtttcaattttttttaaataaaaaaaattcatttttttgaaatgggaaaaaatgc 67
DB 1 GTTTCATATTTTTCATAAATAAAAAAATTCATTTTTTTTGAATGGAAAAATGC 60
QY 68 taaggactaagtgtgtgtattagttattgtttattgattttttttctcagtgga 127
DB 61 TAAGGACTAATGTGTGTATTAGTATTGTGTTATTGGATTTTTTCTTCAGTGA 120
QY 128 aagctagtgtttcttatgatgcagagctataatcatataaatgggaaaaaaattctta 187
DB 121 AAGCTAGTGTCTTATGATGACAGAGCTATATCATATAATGGAAAGAAAAATCTTA 180
QY 188 ttcttggtcaattcattcaagaagcactccacagatggtgcctgattctatacaaa 247
DB 181 TTCTGGTTCAATTCATATFCCAAGAGCACCACAGATGTGGCTGATCTTATACAA 240
QY 248 aggcataaagatgagagcttagatgttatgaaacttatgttttctggaatggacatgagc 307
DB 241 AGGCTAAGATGGAGGCTTAGATGTTATGCAACTTATGTTTCTGGAATGGACATGAGC 300
QY 308 ctctcctggaataataattttgaaggagatgatgatctgttagatcacaataatgg 367
DB 301 CTCTCTCTGGAATAATAATTTTGAAGGAGATATGATCTTGTAGATCATCAAAATGG 360
QY 368 tacaagagcaggaacttatgtcaatttcacgtattggcccttacgtctgctggaatgga 427
DB 361 TACAAAGACGAGGACCTTATGTCAATTTACGTATTGGCCCTTACGTCGTGTGAAATGA 420
QY 428 actttggggattccctgtttggctaaaaatagtgcctggttatggaattgaacaaca 487
DB 421 ACTTTGGGGATTCCTGTGTGCTAAATATGCTGCTGTGTAATTTAGAACAAACA 480
QY 488 atcagccttttaagtggtgtatgaaggatttggttcagaaaaatagtcacatgatgaagt 547
DB 481 ATCAGCCCTTTAAGGTGGCTATGCAAGGATTTGTTCCAGAAAAATAGTCAACATGATGAAGT 540
QY 548 cagaaaattgtttgatctcaaggaggaccaataattatggccagatacaaaaatgaat 607
DB 541 CAGAAAATTTGTTGAATCTCAAGGAGGACCAATAATATTGCCCCAGATAGAAATGAGT 600

QY 608 atggaccagtagaataatgggaaat 629
DB 601 ATGGACCAGTAGAATGGAAAT 622

RESULT 8

BM410481

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

tomato.

Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; euasterids I; Solanales; Solanaceae; Solanum;

Lycopersicon.

1 (bases 1 to 620)

REFERENCE

AUTHORS

Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai

,J., Bougri,O., Kirkness,E., Utterback,T., Van Aken,S., Renning

,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.

Generation of ESTs from tomato fruit tissue, breaker stage (2002)

Unpublished (2002)

Contact: CUGI

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Email: <http://www.genome.clemson.edu/orders/index.html>

This clone is available through the Clemson University Genomics

Institute

Seq primer: T3

Location/Qualifiers

1. .620

/organism="Lycopersicon esculentum"

/cultivar="TA496"

/db_xref="taxon:4081"

/clone_lib="cLEG53G24"

/clone_lib="tomato breaker fruit"

/tissue_type="Pericarp"

/dev_stage="breaker"

/lab_host="SOLR"

/note="Vector: pBluescriptSKmCvAdapt; Site_1: EcoRI;

Site_2: XhoI; supplier: Boyce Thompson Institute;

sequencing: The Institute for Genomic Research. Fruit

were harvested at the breaker stage (first sign of

lycopene accumulation on the blossom end of fruit). Fruit

were cut in half and the seeds and locules were discarded

prior to freezing the pericarp."

188 a 128 c 141 g 163 t

BASE COUNT

ORIGIN

Query Match 24.0%; Score 613.6; DB 10; Length 620;

Best Local Similarity 99.4%; Pred. No. 2.9e-76;

Matches 616; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1210 gtaaaagtacatttcagaataggccatacaatctgcctccatggtccatcagcattctt 1269

DB 1 GTAAAAAGTCACCTTTTCAAGATAGGCCATCAATCTCCCTCCATGGTCCATCAGCATCTT 60

QY 1270 cccgactgcaaaactgccgtttacacactgcacaggttaactctcaagctcgagcata 1329

DB 61 CCCGACTGCAAAACTGCCGTTTACACACTGCACAGCTTAACCTCAAGCTCGACATA 120

QY 1330 aagatgacgcctgcaggtggtgattctcttggcagctacatacaatgaagaacgcctact 1389

DB 121 AAGATGACCCCTGCAGGTGGTGTGTTCTTGTGCACTATACAAATGAAGAAACGCCTACT 180

QY 1390 gctgatgacagcgatacacttacagctaacgactatggagactatggagacagaaaaacgtcacaga 1449

Db 181 GCTGATGACGATACACTTACAGCTAACGACATATGGGAACAGAAAAAGCTCAAGA 240
 QY 1450 gattcatcagactatctgtgtacatgacaaatgtaaatatagcatcttaataagattt 1509
 Db 241 GATTTCATCAGACTATCTGTGGTACATGACAAATGTAATATAGCATCTAATGAAGATT 300
 QY 1510 ctaagaacggaagatctcttatctcactgttatgtccgtgtcgtcatgttctgtt 1569
 Db 301 CTAAGAAACGGAAGATGCTTATCTACTGTATGTCGCTGGTCAATGCTTGCATGTT 360
 QY 1570 ttcgtcaatgaaactatcaggaactgtttatgtatcatatgataatccaaacttaca 1629
 Db 361 TTCGTAATGGAACACTATCAGAACTGTTATGTATGATCATTTGATAATCCAAACTTACA 420
 QY 1630 tacagtggcaactgaaagttaagaactgttattacaagattctctcgtcagttttcc 1689
 Db 421 TACAGTGGCAACGCTGAAGCTTAAGAGCTGGTATTAAACAAGATTCTCTGCTCAGTCTTCC 480
 QY 1690 gttggtctccgaaacgttggcgtgcattatgatacatggaatgcaggagttctaggtcca 1749
 Db 481 GTTGTCTCCGAAACGTTGGCGTGCATTATGATACATGGAATGCAGAGATTCTAGTGCCA 540
 QY 1750 gtcagttgacggtctcaatgaagggtcaagaacttggcgaacacaaatggtcttac 1809
 Db 541 GTCAGTTGAGCGGTCTCAATGAAGGGTCAAGAACTTTGGCGAAACAGAAATGGTCTTTA 600
 QY 1810 aaggttggtctgaaggcga 1829
 Db 601 CAGGTTGCTGAAGGCGA 620

RESULT 9
 LOCUS AW223645 596 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST300456 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA clone cLEM12J23, mRNA sequence.
 ACCESSION AW223645
 VERSION AW223645.1 GI:6535329
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 596)
 Alcala, J., Vrebalov, J., White, R., Matern, A.L., Holt, I.E., Liang, F., Upton, J., Hansen, T., Craven, M.B., Bowman, C.L., Ahn, S., Ronning, C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from tomato fruit tissue
 Unpublished (1999)
 Contact: CUGI
 CLEMSON UNIVERSITY Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 5 prime sequence.
 FEATURES
 Location/Qualifiers
 1..596
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cLEM12J23"
 /clone_lib="tomato fruit red ripe, TAMU"
 /tissue_type="pericarp"
 /dev_stages="red ripe (7-20 days post-breaker)"
 /note="vector: pBluescript SK(-); Site1: EcoRI; Site2: XhoI; supplier: Giovannoni; Fruit were tagged at the breaker stage (first sign of lycopene accumulation on the blossom end of the fruit) and harvested 7 days post-breaker (fully red-ripe), 10 days post breaker, and 20 days post-breaker (over-ripe). 20 day fruit which showed external or internal signs of pathogenesis were

discarded. Fruit were cut in half and the seeds and locules were discarded prior to freezing the pericarp."

BASE COUNT 169 a 113 c 157 g 157 t
 ORIGIN

Query Match 23.0%; Score 586.4; DB 9; Length 596;
 Best Local Similarity 99.0%; Pred. No. 1.8e-72;
 Matches 590; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1513 aagaacggaagatccttatctcactgttatgtccgtgtcgtcatgttcttgcatttttc 1572
 Db 1 AAGAACGGAAGATGCTTATCTACTGTATTATGTCGCTGGTCAATGCTTGCATGTTTC 60
 QY 1573 gccaatggaactatcaggaactgtttatgtatcatatgataatccaaacttacaac 1632
 Db 61 GTCATGGAACACTATCAGAACTGTTTATGTATGGAATTAATCCAAACTTACATAC 120
 QY 1633 agtggcaactgaaagttaagaactgttattacaagattctctcgtcagttttccgtt 1692
 Db 121 AGTGGCAACGCTGAAGTTAAGAGCTGGTATTAAACAAGATTCTCTGCTCAGTCTTCCGTT 180
 QY 1693 ggtctccgaaacgttggcgtgcattatgatacatggaatgcaggagttctaggtccagtc 1752
 Db 181 GGTCTCCGAAACGTTGGCGTGCATTATGATACATGGAATGCAGAGATTCTAGTCCAGTC 240
 QY 1753 acgttgagcgtctcaatgaagggtcaagaacttggcgaacacaaatggtcttacaag 1812
 Db 241 ACGTTGAGCGGTCTCAATGAAGGGTCAAGAACTTTGGCGAAACAGAAATGGTCTTACAG 300
 QY 1813 gttggtctgaaggcgaactcgttaagtctcactccttaagtggaggttcttctgtttaa 1872
 Db 301 GTTGTCTGAAGGCGAATCGTTAAGTCTTCACTCTTAAAGTGGAGTTCTTCTGTTGAA 360
 QY 1873 tgggttcaggttctcactaatggctcaaaagcggccctgacttggtaagcgtcacatt 1932
 Db 361 TGGGTTCGAGGTCTACTAGTGGCTCAAAAGCAGCCCTGACTTGTGTACAAAGCTACATT 420
 QY 1933 aacgcgcctggaggaatgatccactagcttttagacatggcgaagtatgggaaagtcag 1992
 Db 421 AACGGCCTGGAGGAATGATCCACTAGCTTTAGACATGGCAAGTATGGGAAAAAGTCTAG 480
 QY 1993 atattgataaatgttgaaggcgttagtgcctcattggcctggatatacatgacgaagcgcac 2052
 Db 481 ATATGGATAAATGGTGAAGCGTAGTGCCTGCCATTGGCTGGATACATAGCACAAAGCGAC 540
 QY 2053 tgcacaaatgcagttatgtcgtgaacgttcaacgagaagaagtgcggaactaactg 2108
 Db 541 TGCAGCAATGCAGTTATGCTGGAAACGTTCAACGAGGAGAGTGCAGACTAACTG 596

RESULT 10
 LOCUS AW092855 589 bp mRNA linear EST 18-MAY-2001
 DEFINITION EST286035 tomato mixed elicitor, BTI Lycopersicon esculentum cDNA clone cLET22F1, mRNA sequence.
 ACCESSION AW092855
 VERSION AW092855.1 GI:6058450
 KEYWORDS EST.
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 1 (bases 1 to 589)
 D'Ascenzo, M., He, X., Lyman, J., Holt, I.E., Liang, F., Upton, J., Ronning, C.M., Craven, M.B., Fujii, C.Y., Bowman, C.L., Nierman, W., Fraser, C.M., Venter, J.C., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
 Generation of ESTs from tomato leaf tissue
 Unpublished (1999)
 Contact: CUGI

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
Source

Location/Qualifiers

1..589
/organism="Lycopersicon esculentum"
/cultivar="Rio Grande PtoR"
/db_xref="taxon:4081"
/clone="cLEP22F1"
/clone_lib="tomato mixed elicitor, BTI"
/tissue_type="leaf"
/dev_stage="4-6 week old plants"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; cLEP - Inoculated with a variety of disease response
elicitors. Plants exposed to 2,6 dichloroisonicotinic
acid, BTH, jasmonic acid, ethylene, fenthion, EIX,
okadaic acid, or systemin prior to tissue harvest. EcoRI
site was destroyed during cloning."
194 a 102 c 143 g 150 t

BASE COUNT 194 a 102 c 143 g 150 t
ORIGIN
Query Match 22.7%; Score 581; DB 9; Length 589;
Best Local Similarity 99.2%; Pred. No. 1e-71;
Matches 584; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1833 gtttaagtcttcaactcctaagtggaggtctctctgtgaattgggttcaggttcactaat 1892
Db 1 GTTAAGCTCTCACTCTTAAGTGGGAGTCTTCTGTGTGAATGGGTTCAGGTTCACTAGT 60
QY 1893 ggtctaaagcagccctgacttggtacaaggtcatttaacgcccgtggagaaatga 1952
Db 61 GCGTCAAAAGCAGCCCTGACTTGTTGACAGGCTACATTTACGCGCTTGAGGAATGA 120
QY 1953 tccactagcttttagacatggcaagtatgggaaaggtcagatgatgataaatggtagag 2012
Db 121 TCCACTAGCTTTAGACATGGCAAGTATGGGAAAGTCAGATATGGATAAATGGTGAAGG 180
QY 2013 ctataggtccattggcttgatacagacagaagcgactcagcaaatcagattatgc 2072
Db 181 CGTAGTCCCATTTGGCTTGGATACATAGCACAAAGCGACTGCAGCAATGCAGTTATGC 240
QY 2073 tggaacgttcaacgagaagaagtgcggaactaactcgggacacaccttccagagatgga 2132
Db 241 TGGAGGTTCAAGAGAGAAAGTGGCAGACTTAACCTGCGGACAACTTCTCAGAGATGTA 300
QY 2133 ccattgtccacgatcgtgctgaaacccaagtggaaaccttgttagtagtatttcgaagaatg 2192
Db 301 CCATGTTCCACGATCGTGGCTGAAACCAAGTGGAACTTGTGTAGTAGTATTTCGAAGAATG 360
QY 2193 gggaggttaaccacagaattctctagtcaggatcaagaataaagaactcgaagaagt 2252
Db 361 GGGAGGTAACTCAACAGGAAATTTCTAGTCAGGAGATCAAGATAAAGAACTCGAAAAAGT 420
QY 2253 aaaaactgttcagtaactatggtcgttggaattcgcgcgaaataacatacacagaagcta 2312
Db 421 AAAAAGTGTTCAGTAACATATGCTTGAATTCGCGCGCAAAATACATACAGGAAGTA 480
QY 2313 acaatggagggtcacgtttgcgaattgcagctggaataaacaattagaagataaagaata 2372
Db 481 ACAATGGAGGCTACAGTTTGCAGATTCAGCTGAAATAAACCAATTAGAAGATAAAGAATA 540
QY 2373 ttgtattaaaaggagatataaattacagagaattttctttattcttt 2421
Db 541 TTTGATTAAGAGAGTGTATAAATTTACAGAGAAATTTCTTTATTCTTT 589

RESULT 11
AW032987
LOCUS
563 bp mRNA linear EST 18-MAY-2001

DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

EST276546 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC17L14, mRNA sequence.
AW032987
AW032987.1 GI:5891743
EST
tomato.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon
1 (bases 1 to 563)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
Location/Qualifiers
1..563
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC17L14"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBlueScript SK(-); Site.1: EcoRI; Site.2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination with no selection, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
159 a 106 c 135 g 163 t

FEATURES
source

Query Match 21.8%; Score 556.6; DB 9; Length 563;
Best Local Similarity 99.3%; Pred. No. 2.6e-68;
Matches 559; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 616 gtagaatgggaaattggtgctctgtgtaaaagcttatatacaaaatggcgagctcaaatgct 675
Db 1 GTAGAATGGGAAATTTGGTCTCTCTGGGAAGCTTATACAAAATGGCAGCTCAAAATGGCT 60
QY 676 gtagggttgaaaaactggtgccatggatcatgtgaagcaagagatgctctgatct 735
Db 61 GTAGGTTTGAANAACCTGGTGTCCCATGGATCATGTGTAGCAAGAGATGCTCTGATCT 120
QY 736 gtgattgatacttgaatggcttctactgcgaaggggttcctcctaataaagccttcaaaa 795
Db 121 GTGATTGATCTGTAAATGGCTTCTACTGCGAAGGTTCCGCTCCCTAATAAGCTTACAAA 180
QY 796 cctaaaaatgtggacagaagatgactggtgtgtatagaaaattcgggtgggtcccaattct 855
Db 181 CCTAAAATGTGGACAGAAAGTATGGACTGGCTGGTATACGAAATTCGGTGGTCCAAATCT 240
QY 856 caaagcagccgaagacattgcatttcagttgcagggtttgttcagaaacaatggttca 915
Db 241 CAAAGCAGCGGAAGACATTCATTTTCAGTTGCCAGGTTGTTCATAACAAATGGTTCA 300
QY 916 ttcttcaattactacatgtatcatatggaggaaacaattttggccggacatcatcagggtt 975
Db 301 TTCTTCAATTACTACATGTATCATGGAGAACAAAATTTTGGCCGGACATCATCAGGGCTT 360
QY 976 ttcatgcaactagtagatttatgatctctctcgatgaatatggttctgctgaatgaa 1035
|||||

```

Db 361 TTCATTGCAACTAGCTACGATTATGATGCTCTCTCGATGAATATATGGGTGCTGAATGAA 420
QY 1036 ccaaaagtatggcacttgagagacttacataaagctatcaagctatctgaaccgcttta 1095
Db 421 CCAAAGTAGTGGCAGCTTGAGAGACTTACATAAAGCTATCAAGCTATCTGAACCGGCTTTA 480
QY 1096 gtttcacatagtctgcggctgactagtcttggaaagtaatacaagagctcatgtttataga 1155
Db 481 GTTTCATCATATGCTCGCGTGACTAGCTTGTGAAGTAAACAAGAGGCTCATGTGTTATAGA 540
QY 1156 tcaaaatctggagcttgctgctgc 1178
Db 541 TCAAAATCTGGAGCTTGCTGCTC 563

RESULT 12
BE431542
LOCUS BE431542 551 bp mRNA linear EST 18-MAY-2001
DEFINITION EST336357 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
clone cLEG6017, mRNA sequence.
ACCESSION BE431542
VERSION BE431542.1 GI:9429385
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 551)
AUTHORS Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
,S.D.
TITLE Generation of ESTs from tomato fruit tissue, breaker stage
JOURNAL Unpublished (2000)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

FEATURES
Location/Qualifiers
source 1..551
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEG6017"
/clone_lib="tomato breaker fruit, TIGR"
/tissue_type="pericarp"
/dev_stage="breaker"
/lab_host="SOLR"
/note="Vector: pBluescriptSKmCudapt; Site_1: EcoRI;
Site_2: XhoI; Fruit were harvested at the breaker stage
(first sign of lycopene accumulation on the blossom end of
the fruit). Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

BASE COUNT 153 a 120 c 116 g 162 t
ORIGIN

Query Match 21.5%; Score 549.4; DB 10; Length 551;
Best Local Similarity 99.8%; Pred. No. 2.6e-67;
Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 819 gactggctgtatcagaattcgggtgttcgaattctcctcaagaccgacgaagacattgc 878
Db 1 GACTGGCTGTGTATCAGAAATTCGGTGTCCAAATTCCTCAAGACCAGCGGAAGACATTGC 60
QY 879 atttcagttgcagggttggttcagaacaattggttcattctcaattactacatgtatca 938
Db 61 ATTTTCAGTTGCCAGGTTGTTTCAGAACAAATGGTTCATTCTTCAATTACTACATGTATCA 120

```

```

QY 939 tggaggacaacattttggcggacatcatcagggtttttcattgcaactagctacgattca 998
Db 121 TGGAGGAACAAATTTTGGCCGGACATCATCAGGCCTTTTCATTGCAACTAGCTACGATTA 180
QY 999 tgatgctctcgtatgaatatgggttgctgaatgaacaaagatgaggaacttgagaga 1058
Db 181 TGATGCTCTCTCGATGAATATGGGTGCTGAATGAACCAAAAGTATGGGCACCTTGAGAGA 240
QY 1059 ctacataaagctatacaagctatctgaaccgcttttagtttcatcatatgctgcgctgac 1118
Db 241 CTTACATAAAGCTATCAAGCTATCTGAACCGGCTTTAGTTTTCATATATGCTGCGGTGAC 300
QY 1119 tagtcttggaaagtaatacaagagctcatgtttatagatcaaaatctggagcttgctgc 1178
Db 301 TAGTCTTGGAAAGTAATCAAGAGGCTCATGTTATAGATCAAAATCTGGAGCTTGCTGTC 360
QY 1179 tttttatccaactatgactctagatattcgaataaagtaacaccttcagaataggccata 1238
Db 361 TTTTATCCAACTATGACTCTAGATATTTCAGTAAAGTACACCTTTTCAGATAGGCCATA 420
QY 1239 caatctgctccatgctccatcagcattctcccgactcaaaactgccgtttacaacac 1298
Db 421 CAATCTGCCCTCATGGTCCATCAGCAATCTTCCGACTCAAAACTGCCGTTTACAAAC 480
QY 1299 tgcacaggttaactctcaagctcggagcagcagcagcagcagcagcagcagcagcagc 1358
Db 481 TGCACAGGTAACTCTCAAAAGCTCGAGCATAAAGATGACGCTGACAGGTGGTGGATTGTC 540
QY 1359 ttggcagtcacat 1369
Db 541 TTGGCAGTCAT 551

RESULT 13
AW033711
LOCUS AW033711 549 bp mRNA linear EST 18-MAY-2001
DEFINITION EST277282 tomato callus, TAMU Lycopersicon esculentum cDNA clone
cLEC29B5 similar to beta-galactosidase, mRNA sequence.
ACCESSION AW033711
VERSION AW033711.1 GI:5892467
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
REFERENCE 1 (bases 1 to 549)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.

FEATURES
Location/Qualifiers
source 1..549
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC29B5"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.

```


SOURCE
ORGANISM
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 611)
Alcala, J., Vrebalov, J., White, R., Matern, A. L., Vision, T., Holt, I. E.,
Liang, F., Upton, J., Craven, M. B., Bowman, C. L., Ahn, S., Ronning,
C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.

FEATURES
source
1. 611
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cJEC6M3"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XLI-Blue MRF"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
XhoI; supplier: Giovannoni laboratory; cLEC - Cotyledons
of seedlings 7-10 days post-germination were excised, cut
at both ends and placed on MS medium with no selection.
Mixed callus was harvested at 25 and 40 days and included
undifferentiated masses. Tomato Callus EST Library"
186 a 104 c 144 g 177 t

Query Match 20.8%; Score 532.4; DB 9; Length 611;
Best Local Similarity 90.0%; Pred. No. 5.6e-65;
Matches 610; Conservative 0; Mismatches 1; Indels 67; Gaps 1;
Qy 302 atgagccttctctggaataataatttgaggagagatgatctgttagattcatca 361
Db 1 ATGAGCCCTTCCTCGGAAATATATATTTTGAAGGAAGATATGATCTGTGTAGATTATCA 60
Qy 362 aaatggtacaagagcagcactttatgtcaatttcacgtattgccccttacgtctgtctg 421
Db 61 AAATGGTACAAAGAGCAGGACTTTATGTCAATTTACGTATTGGCCCTTACGCTCTGTGCTG 120
Qy 422 aatggaacttgggggattccctgtttggctaaaataatgtgcctggtatggaatttagaa 481
Db 121 AATGGAACCTTG----- 132
Qy 482 caacaatcagccttttaaggtggctatgcaaggattttgttcagaaaaatagttcaacatga 541
Db 133 -----GGTGGCTATGCAAGGATTTGTTTCAGAAATAAGTCAACATGA 173
Qy 542 tgaagtcagaaaatttggtaaatcgaaggaggaccataaattatgccccagatcacaaa 601
Db 174 TGAGTCAGAAAATTTGTTGTAATTCAGAGGAGGACCAATAATATTATGGCCCATAGATAAA 233
Qy 602 atgagatggaccagtagaattgggaattgtgctccctggttaagcttatcaaaaatggg 661
Db 234 ATGAGTATGGACCATAGAAATGGGAAATGGTGTCTCTGGTAAAGCTTATACAAAATGGG 293
Qy 662 cagctcaaatggctgtaggtttgaaaactgggtgtcccatggatcatgtgtgaagcaagagg 721
Db 294 CAGCTCAATGGCTGTAGGTTTGAAACTGGTGTCCCATGGATCATGTGTAAAGCAAGAGG 353
Qy 722 atgctcctgatcctgtgattgatacttgtaattggtcttactgcgaaggggttccgtccta 781
Db 354 ATGCTCCCTGATCCTGTGATTGATATTGTAATGGCTTCTACTGCGAAGGGTTCCTGCTTA 413

Qy 782 ataaqccttataaaacctaataatgtgacagaagatgtgactgctggtgatatcaaaattcg 841
Db 414 ATAAAGCCTTACAAACCTTAATATGTGGACAGAAAGTATGGACTGGTATGATACGAATTCG 473
Qy 842 gtggtccaatctctcaaaagaccagccgaagacattgcattttcagttgccaggtttgttc 901
Db 474 GTGGTCCAATTCCTCAAAAGACCAGCCGAAGACATTGCATTTTCAGTTGCCAGGTTGTTC 533
Qy 902 agaacaatgggttcattcttcaattactactatcatgtatcatggaggaacaaattttggccgga 961
Db 534 AGAACAAATGGTTTCATTTCTCAATTACTATCATGTATGAGGAGCAAAATTTTGGCCGGA 593
Qy 962 catcatcagggccttttca 979
Db 594 CATCATCAGGGCTTTTCA 611
Search completed: June 25, 2002, 04:37:44
Job time: 7739 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 02:29:50 ; Search time 92.36 Seconds
(without alignments)
6792.421 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2534
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1205.8	47.2	2628	2	US-08-696-944-1
2	1171.2	45.9	2944	2	US-08-696-944-18
3	1168.2	45.7	2945	2	US-08-687-372-2
4	67	2.6	2322	3	US-09-436-605-1
C 5	54.8	2.1	7218	1	US-08-232-463-14
C 6	54.4	2.1	8920	2	US-08-446-855A-1
C 7	54.4	2.1	8920	4	US-09-150-741-1
8	48.8	1.9	19124	2	US-08-487-826B-13
9	47.4	1.9	3057	1	US-08-551-459-3
10	46.8	1.8	1493	6	5340934-5
11	46.2	1.8	3871	2	US-08-599-455B-3
C 12	46.2	1.8	3871	4	US-09-069-781B-3
C 13	45.6	1.8	470	4	US-09-020-956-102
C 14	45.6	1.8	470	4	US-09-030-607-102
C 15	45.6	1.8	470	4	US-09-439-313-102
16	45.2	1.8	3515	1	US-08-596-985-1
17	45	1.8	1641	1	US-08-300-903A-8
18	45	1.8	2445	6	5215909-9
19	44.4	1.7	1160	3	US-08-995-159-1
20	43.8	1.7	2110	4	US-09-419-459-1
21	43.8	1.7	3466	1	US-08-551-459-5
22	43.4	1.7	1240	1	US-08-248-466B-6
23	43.4	1.7	1510	1	US-08-248-466B-3
C 24	43.4	1.7	5852	1	US-07-867-106-2
25	43.4	1.7	8920	2	US-08-446-855A-1
26	43.4	1.7	8920	4	US-09-150-741-1
C 27	43.2	1.7	509	4	US-09-030-607-202

C 28	43.2	1.7	509	4	US-09-439-313-202	Sequence 202, Appl
C 29	43	1.7	1431	4	US-09-316-083-2	Sequence 2, Appli
C 30	43	1.7	3527	2	US-08-909-965C-7	Sequence 7, Appli
31	42.4	1.7	1172	1	US-07-945-288-9	Sequence 9, Appli
32	42.4	1.7	1172	1	US-08-462-831-9	Sequence 9, Appli
33	42.4	1.7	1172	1	US-08-461-809-9	Sequence 9, Appli
34	42.4	1.7	1172	1	US-08-461-441-9	Sequence 9, Appli
35	42.4	1.7	1172	5	PCT-US93-08518-9	Sequence 9, Appli
36	42.4	1.7	2550	6	5258287-23	Patent No. 5258287
37	42.4	1.7	6243	2	US-09-056-075-1	Sequence 1, Appli
C 38	42.4	1.7	19124	2	US-08-487-826B-13	Sequence 13, Appli
C 39	42.2	1.7	857	1	US-08-308-883-1	Sequence 1, Appli
C 40	42.2	1.7	857	1	US-08-730-163-1	Sequence 1, Appli
C 41	42.2	1.7	857	4	US-08-256-799-1	Sequence 1, Appli
C 42	42.2	1.7	857	4	US-08-462-437-1	Sequence 1, Appli
43	42.2	1.7	2058	2	US-08-749-391-1	Sequence 1, Appli
44	42.2	1.7	2058	3	US-09-390-200-1	Sequence 1, Appli
45	42.2	1.7	2322	1	US-08-618-164-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-696-944-1
; Sequence 1, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2628 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 130..2319
; US-08-696-944-1

Query Match 47.2%; Score 1205.8; DB 2; Length 2628;
Best Local Similarity 73.5%; Pred. No. 2.9e-268;
Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;

Qy	98	tttggttattggatttttttcttcctcagtgtaaaagctagtggttcttccttatgatcagacagacta	157
Db	194	TGTTATTGTTGTTTTTGGGGTTTGTATTGTCACAGCCTCTGTTACTTATGATCATAAAGCCA	253
Qy	158	taatcaataatgggaaagaaaaattcttatcttattcttggttcaattcattatccaagaagca	217
Db	254	TTATGATTAAATGGCAGAGAAATTTTGATCTCTGGTTCCCATTCACATCCAGAAGACA	313
Qy	218	ctccacagatgtagcctgatctttatacaaaaagctaaagatggagccttagatgtattg	277
Db	314	CACCTCAGATGTGCCAGACCTTATTCAAAAGGCCAAAGATGAGGGCTTGATGTTATAG	373
Qy	278	aaacttatgttttctggaaatggacatgagcctctctctcctggaaaaataataatttgaagaa	337
Db	374	AGACTTATGTGTTCTGGAATGGACATGAACCTTCTCTGGAAAAATATTATTTTGAGGATA	433
Qy	338	gatatgatctgttagatcatcaaaaatggtacaaaagcaggaaccttatgtatcaatttac	397
Db	434	GGTTTGACCTGTGTGGGTTTCATAAAAGTTGGTTCAGCAGCTGGTCTATTGTTTATCTCA	493
Qy	398	gtattggcccttaagctctgtctgctgaatgaaactttgggggattccctgtttggctaaaaa	457
Db	494	GGATTGGCTCTTCATATGTGCTGATGAAGAACTTTGGAGGATTTCCTGTTTGGCTCAAT	553
Qy	458	atgtgcctcgttatggaattttagaacaacaaatcaagccttttaagtggtgctatgcaagat	517
Db	554	ATGTTCTCTGGTATTGCTTTTCAGAACAGACAAATGAGCTTTCAAGGAGGCAATGCAAAAAT	613
Qy	518	ttgttcgaaaaatagtcacaatgatgaagtcaaaaaatttttgaaatctcaagagagac	577
Db	614	TCACTGGAAGAGATTGTAATATTAATGAAGACAGAGAAGTTGTTTCAATCCCGAGGAGTC	673
Qy	578	caataattatggccagatatacaaaatgagtatggaccagtagaaatgggaaattggtgtc	637
Db	674	CAATAAATCTGCTCAGATAGAGAAATGACTATGACCACTGGAATGGGAAATTTGGTCTC	733
Qy	638	ctggtaagccttatcaaaaatgggcagctcaaaatggctgttaggtttgaaaactgggtgcc	697
Db	734	CTGAAAAGCTTATACCAAATGGGCTGCTCAATGGCTGTAGGTCTAGATACTGGTGTC	793
Qy	698	catgatcatgtataagcaagagatgctccctgatcccttgattgattgatacttgtaagct	757
Db	794	CATGGGTTATGTCCAGCAAGAAGATGCACCTTGATCCCTATTATTGTGATACCTGCAATGGAT	853
Qy	758	tctactcggaagggttccgtccctaataagccttacaacccataaaatgtggacagaagat	817
Db	854	TTTACTGTGAAAACTTCACTCCAAAACAAGAACTACAAACCCAAAATTTGGGACAGAAAAAT	913
Qy	818	ggactggctggtatagaataatcggttggtccaatctctcaagaaccagccgaagaacattg	877
Db	914	GGACTGGCTGGTACACTGCTTTTGGTGGTGAACCCCTTATAGACCAGCAGAAGATATAG	973
Qy	878	cattttcaagtgtccaggtttgttcagaacaaatggttcattcttcaatcactacatgctc	937
Db	974	CATTTTCAGTTGCCAGATTCATTTCAGATCGCGGCTCTACTCTTTAACTACTATATGATATC	1033
Qy	938	atggaggaacaaattttggccggacatcatcagggcttttcattgcaactagctacgatt	997
Db	1034	ATGGAGGAACTAATTTTGGCCGGACATCGAATGGCTCTCTGTTGCCACAAGTTATGACT	1093
Qy	998	atgatgctcctctcgatgaatatgggttgctgaatgaacccaaaagtaatgggcacttgagag	1057
Db	1094	ATGATGCTCCCATTTGATGAATATFGACTTCTTAATTAACCAAAAAATGGGGGCACTGAGAG	1153
Qy	1058	acttacataaagctataagctatctgaaccggccttttagtttcaatcatatgctcgggtga	1117
Db	1154	AATTAACATAGAGCAATAAACAATCGGAGTCGGCTTTAGTGTGGTGCATCCACAGTGT	1213
Qy	1118	ctagctcttgggaagtaatcaagaggtctatgtttatagatcaaaatcttggagcttgtgctg	1177
Db	1214	CATGGCCTGAAAAAACCTTTGAGGTACATTTTGTACAAAGACAGAGTCT---GCCGTGCTG	1270

6

;

Query Match 2.6%; Score 67; DB 3; Length 2322;
Best Local Similarity 49.2%; Pred. NO. 1.5e-06;
Matches 209; Conservative 0; Mismatches 210; Indels 6; Gaps 1;

Qy 187 atttctgttcaattcattatccagaagcactccacagatggtgcctgatctttatacaa 246
Db 175 atttctgttcaattcattatccagaagcactccacagatggtgcctgatctttatacaa 234
Qy 247 aaggctaaagatggaggtctagatgtattgaaccttatgttttctggaatggacatgag 306
Db 235 aagatgaagatggtggctgaatccatccacagacgtgacctggaactttcacgaa 294
Qy 307 ccttctctggaataataatttgaagaagatgatcttctgttagattcatcaaaatg 366
Db 295 ccccgagcgagacagatccagtttctctgggagcaggtggaatatttttaagctg 354
Qy 367 gtacaaagacgagcatttatgcataattacgtatgtgccccttaagctgtgctgaatgg 426
Db 355 gcccatgagctgggactgctgctatcctgagcgggacctatattctgtgcagagtgg 414
Qy 427 aacttgggggattccctgttggctaaaatgtgcctggtatggaatttagaacaac 486
Db 415 gacatgggaggtattacctgttctgttatttataaagaatctattctcgttctct 474
Qy 487 aatcagcctttaaagtggtgctatgcaaggatttgttcagaaaatagtcacacatgatgaag 546
Db 475 gatccagattacctgtgagctgtgacaaatggtggagtcctcctgcccaagatgaag 534
Qy 547 tcagaaaattgttgaatctcaagagagaccataaattatggtgcccagatcaacaataatgag 606
Db 535 cc-----tctctctatcagaacgagggcgattataaacatgaggttgaaatgaa 588
Qy 607 tatgg 611
Db 589 tatgg 593

RESULT 5
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTgpt-F1s
US-08-232-463-14

Query Match 2.1%; Score 54.8; DB 1; Length 7218;
Best Local Similarity 5.6%; Pred. NO. 0.0014;
Matches 23; Conservative 220; Mismatches 167; Indels 0; Gaps 0;

Qy 1954 ccactagcttttagacatggcaagtatgggaaaaggtcgagatgataaaatggtgaagc 2013
Db 1466 CAAGTAGTTAAAGAGATAGAGAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRR 1407
Qy 2014 gtaggctccattgacctggtgatacagacaaagcgactgcagcaaatgcagttatgct 2073
Db 1406 RRR 1347
Qy 2074 ggaacgttcaacgagaagaagtcgcgaaactaactgcgacacacctctcacagatgttac 2133
Db 1346 RRR 1287
Qy 2134 catgttccacgactggtgctgaaacaaagtgaaactgttagtagtagtattcgaaatgg 2193
Db 1286 RRR 1227
Qy 2194 ggaggtaatccaaacaggaattctctctagtcaggagatcaagataaagaactcgaaaagta 2253
Db 1226 RRR 1167
Qy 2254 aaactgttcagtaactatgctgctgaattcgccgcgcaaaaatacacacgaagctaa 2313
Db 1166 RRR 1107
Qy 2314 caatggaggctacagtttgcaaatgctgagctgaataaacaattagaaatg 2363
Db 1106 RRR 1057

RESULT 6
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A


```

: FILING DATE: 06-Jul-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mithard, Leonard C
: REGISTRATION NUMBER: 29.6
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-816-4000
: TELEFAX: 703-816-4100
: INFORMATION FOR SEQ ID NO: 1
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8920 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: genomic
: US-08-446-855A-1

```

Query Match 2.1%; Score 54.4; DB 2; Length 8920;
Best Local Similarity 49.4%; Pred. No. 0.0018;
Matches 168; Conservative 0; Mismatches 171; Indels 1

Sequence	Matches	Conservative	Mismatches	Indels	Gaps
1	168	0	171	1	1

[illegible]

```

RESULT 7
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Seq
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/0/
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

```

Query Match 2.1%; Score 54.4; DB 4; Length 8920;
Best Local Similarity 49.4%; Pred. NO. 0.0018;
Matches 168; Conservative 0; Mismatches 171; Indels 1

Qy	2212	at	tctctctagtcaggagatacaagataaagaactcgaagaataaaaacttgctcagtaacta	2271
Db	875	AT	ATATATATATCTTTTAAATACAACTTTGTGATGTTAAAGAAATAAAACATGTTTAAAGACCTA	816
Qy	2272	tg	gtgcttggaattcgcgcgcgaaaaaacacacacacgaagcgaacaaatggaggctacagttt	2331
Db	815	TG	ATTCAGACAATATCCCCAATAATTATATATATATATATATATTTATATATATATCTATAT	756
Qy	2332	gcaaat	gcagctgaaataaaacattagaagataaagaaaatatatttgattaaaaggagata	2391
Db	755	TAT	TTTTTTCCTCTTTTCTTTTCTTTTCTTTTATATACATTTATATATATATGTTTAAATATTTA	696
Qy	2392	t	aaatttcacagaga-atttcttctattcttctgaacacttgggtttataaagatttataca	2450
Db	695	TAA	ATTTACATATACAAAGTCATTTTTCATATGTAATAATTTTTTTTTTCTTTTTTTTTT	636
Qy	2451	ga	atttctctgtatttggtattatgagattgagaagattgtacagcttccaaatactatt	2510
Db	635	TTTT	TTTTTTTTTTTTTTTTTAAATAGTAGAATTTACTTTTTTTTTTATAAATCAAGAAAAAAT	576
Qy	2511	aga	tacaaataaatttcattgtaaaaaaataaaaaa	2550
Db	575	AA	ATAATGAATGAATAAATTAATAATAATAATATATATAATAATAA	536

RESULT 8

US-08-487-826B-13
; Sequence 13, Application US/08487826B

; GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellemis, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:

/ CONTACTED: 10/28/83.
 / ADDRESS: Knobbe Martens Olson & Bear
 / STREET: 620 Newport Center Drive 16th Floor
 / CITY: Newport Beach
 / STATE: California
 / COUNTRY: US
 / ZIP: 92660

ZIP: 92000
COMPUTER READABLE FORM.

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israel, N. Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

Db 1482 aaaaaaa 1489

RESULT 11

US-08-599-455B-3

; Sequence 3, Application US/08599455B

; Patent No. 5972621

GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

APPLICANT: Tepper, Robert I.

APPLICANT: Culpepper, Janice A.

TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS THAT

TITLE OF INVENTION: MODULATE BODY WEIGHT USING THE OB RECEPTOR

NUMBER OF SEQUENCES: 44

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/599,455B

FILING DATE: 22-JAN-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/583,153

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: 08/570,142

FILING DATE: 11-DEC-1995

APPLICATION NUMBER: 08/569,485

FILING DATE: 08-DEC-1995

APPLICATION NUMBER: 08/566,622

FILING DATE: 04-DEC-1995

APPLICATION NUMBER: 08/562,663

FILING DATE: 27-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Meiklejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/017001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3871 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

LOCATION: 194....3688

US-08-599-455B-3

Query Match 1.8%; Score 46.2; DB 2; Length 3871;

Best Local Similarity 49.8%; Pred. No. 0.11;

Matches 117; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 2320 aggcacagttgcaaatgacagctgaataaacattagagataaagaataattgttatt 2379

Db 3617 ATGCTCAATCCAACTGTGTCTACTCAGACTCATAAGATCATGGAACCAAGATGTGT 3676

Qy 2380 aaaaggagatataaattacagagaattttcttcttatttcttgtaaaactttggttata 2439

Db 3677 GACCTAACTGTGTAATTTCACTGAAGAAACCTTCAGATTTGTGTATTAATGGGTAATATA 3736

Qy 2440 aagtttatcacagaattttcttcttatttgattatgagattgaagaagattgtacagcttc 2499

Db 3737 AAGTGTATAGATTATAGTTCTGGTGGAGAGAGAAAGAACACAGAGTCCAAATTTGA 3796

Qy 2500 caatactattagatacaataataattctcatgtataaaaaaataaaaaa 2554

Db 3797 AAATAAATGTTCCCAACTGAAAAAATAAAAAAATAAAAAAATAAAAAA 3851

RESULT 12

US-09-069-781B-3

; Sequence 3, Application US/09069781B

; Patent No. 6287782

GENERAL INFORMATION:

APPLICANT: Tartaglia, Louis A.

APPLICANT: Tepper, Robert I.

APPLICANT: Culpepper, Janice A.

APPLICANT: White, David W.

TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR

TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS,

TITLE OF INVENTION: INCLUDING OBESITY AND CACHEXIA

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/069,781B

FILING DATE: 29-APRIL-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/864,564

FILING DATE: 28-MAY-1997

APPLICATION NUMBER: US 08/708,123

FILING DATE: 03-SEP-1996

APPLICATION NUMBER: US 08/638,524

FILING DATE: 26-APR-1996

APPLICATION NUMBER: US 08/599,455

FILING DATE: 22-JAN-1996

APPLICATION NUMBER: US 08/583,153

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: US 08/570,142

FILING DATE: 11-DEC-1995

APPLICATION NUMBER: US 08/569,485

FILING DATE: 08-DEC-1995

APPLICATION NUMBER: US 08/566,622

FILING DATE: 04-DEC-1995

APPLICATION NUMBER: US 08/562,663

FILING DATE: 27-NOV-1995

ATTORNEY/AGENT INFORMATION:

NAME: Meiklejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/082001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3871 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Coding Sequence

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 25, 2002, 03:33:55 ; Search time 489.19 Seconds
(without alignments)
8963.796 Million cell updates/sec

Title: US-09-701-868-4
Perfect score: 2554
Sequence: 1 aaaaaagtttcaattttt.....aaaaaaaaaaaaaaaaaaaaa 2554

Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues 3472872
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : **Genes**
1: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1980.DAT.*
2: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1981.DAT.*
3: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1982.DAT.*
4: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1983.DAT.*
5: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1984.DAT.*
6: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1985.DAT.*
7: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1986.DAT.*
8: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT.*
9: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1988.DAT.*
10: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1989.DAT.*
11: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1990.DAT.*
12: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT.*
13: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1992.DAT.*
14: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1993.DAT.*
15: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT.*
16: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT.*
17: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT.*
18: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT.*
19: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT.*
20: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT.*
21: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT.*
22: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT.*
23: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001B.DAT.*
24: /net/abss06/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2546	99.7	2554	21	AAZ29341 Tomato beta galact
2	1205.8	47.2	2628	16	AAZ01014 Lupin exo-(1-4)bet
3	1172.8	45.9	3069	21	AAZ29340 Tomato beta galact
4	1171.2	45.9	2944	16	AAZ01015 Tomato exo-(1-4)be
5	1168.2	45.7	2945	16	AAQ89602 Tomato beta galact
6	1168.2	45.7	3224	21	AAZ29338 Tomato beta galact
7	1108	43.4	2469	21	AAC49769 Arabidopsis thalia
8	952.4	37.3	2571	21	AAC49810 Arabidopsis thalia
9	877	34.3	2553	21	AAC46270 Arabidopsis thalia

10	877	34.3	2562	21	AAZ29341
11	708.8	27.8	2778	17	AAZ15173
12	649.4	25.4	2972	21	AAZ29344
13	507.8	19.9	2984	21	AAZ29339
14	388.4	15.2	749	21	AAZ29343
15	372.6	14.6	1969	21	AAZ29342
16	331.4	13.0	755	21	AAZ29342
17	242.8	9.5	890	21	AAZ1630
18	231.2	9.1	1240	21	AAZ38606
19	208.6	8.2	458	21	AAZ36838
20	152	6.0	457	21	AAZ44756
21	102.4	4.0	8546	20	AAZ13083
22	96	3.8	8991	20	AAZ13195
23	94	3.7	14736	19	AAV52304
24	91.4	3.6	542	21	AAZ95011
25	86.8	3.4	832	17	AAZ37082
26	76.6	3.0	6279	20	AAZ13280
27	75.6	3.0	513445	22	AAI61373
28	73	2.9	3030	22	AAH42268
29	72.8	2.9	3039	22	AAK94894
30	72.2	2.8	1908	21	AAAG2060
31	72.2	2.8	3024	22	AAK94298
32	72.2	2.8	3106	20	AAZ52245
33	72.2	2.8	3106	22	AAZ54535
34	72.2	2.8	3106	22	AAZ72403
35	72.2	2.8	3234	21	AAAG2070
36	70.6	2.8	819	22	AAK91999
37	70.6	2.8	819	22	AAK93676
38	70.6	2.8	2906	22	AAK94350
39	68.4	2.7	434	22	AAZ52906
40	67	2.6	2322	22	AAZ83033
41	65.2	2.6	440	21	AAZ38419
42	63.8	2.5	1794	20	AAZ88791
43	62.2	2.4	2409	19	AAZ42728
44	62.2	2.4	2435	21	AAZ15878
45	61.8	2.4	669	22	AAZ32666

ALIGNMENTS

RESULT	1
AAZ29341	
ID	AAZ29341 standard; cDNA; 2554 BP.
XX	AC
XX	AAZ29341;
XX	29-FEB-2000 (first entry)
DT	Tomato beta galactosidase-4 cDNA.
XX	
DE	Tomato beta galactosidase-4; TBG-4; clone pZBG2-1-4; fruit ripening;
KW	Rutgers Tomato plant; pectin; biofilm; gene mapping; ds.
XX	Lycopodium esculentum.
OS	
XX	Key Location/Qualifiers
FT	64..2238
FT	/*tag=
FT	/product= "Tomato beta galactosidase-4"
XX	
XX	WO9964564-A1.
PN	
PD	16-DEC-1999.
XX	
PF	08-JUN-1999; 99WO-US12697.
XX	
PR	09-JUN-1998; 98US-0088805.
XX	
PA	(USDA) US DEPT OF AGRICULTURE.
XX	
PI	Gross KC, Smith DL;
XX	

DR WPI: 2000-097532/08.
DR P-PSDB; AAY44306.
XX New beta-galactosidases, used to prepare transgenic plants with altered
PT fruit ripening -
XX
PS Claim 2; Fig 2; 85pp; English.
XX
CC The present sequence encodes tomato beta galactosidase-4 (TBG-4). This
CC is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from
CC breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants.
CC TBG cDNA has been expressed in E.coli strain XlriBlue MR. This is used
CC for modifying cell wall metabolism and controlling ripening of fruit by
CC altering activity of beta galactosidase II protein. Pectin with reduced
CC galactosyl content is produced for use in biofilms or solutions.
CC Fragments of the cDNA are used for gene mapping and for detecting
CC expression of beta galactosidase in plant tissue. Transgenic plants
CC with altered fruit ripening are produced by introducing DNA constructs
CC comprising a TBG cDNA associated with promoter/enhancer elements.
XX
SQ Sequence 2554 BP; 822 A; 431 C; 559 G; 742 T; 0 other;

Query Match 99.7%; Score 2546; DB 21; Length 2554;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2549; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 aaaaaaagttcaatttttttcttaaaaaataaaaaaaattcatttttttgaattgaa 60
Dbb 1 aaaaaaagttcaatttttttcttaaaaaataaaaaaaattcatttttttgaattgaa 60
Qy 61 aaaaagctaaagactaatgtgtgtgtattattgtattttttttttttttttct 120
Dbb 61 aaaaagctaaagactaatgtgtgtgtattattgtattttttttttttttttct 120
Qy 121 tcaagtgaagctagtggttttcttatgatgacagagctataataatcggaagaagaaa 180
Dbb 121 tcaagtgaagctagtggttttcttatgatgacagagctataataatcggaagaagaaa 180
Qy 181 attctattttctgttcattcattatccaaagaagcactccacagatgacctgatctt 240
Dbb 181 attctattttctgttcattcattatccaaagaagcactccacagatgacctgatctt 240
Qy 241 atacaaaggctaaagatggaggcttagatgtattgaaacttatgtttttctggaatgga 300
Dbb 241 atacaaaggctaaagatggaggcttagatgtattgaaacttatgtttttctggaatgga 300
Qy 301 catgagccttctcctggaaaataaattttgaaggaagatgatctctgttagatcactc 360
Dbb 301 catgagccttctcctggaaaataaattttgaaggaagatgatctctgttagatcactc 360
Qy 361 aaaaagctacaaagacagagctttatgtcaattacgtatttgcccttacgctctgtgct 420
Dbb 361 aaaaagctacaaagacagagctttatgtcaattacgtatttgcccttacgctctgtgct 420
Qy 421 gaatggaaactttgggggattccctgttttggctaaaataatgtgctgtggtatgaaatttga 480
Dbb 421 gaatggaaactttgggggattccctgttttggctaaaataatgtgctgtggtatgaaatttga 480
Qy 481 acaaaacaatcagccttttaaggtggctatcaagagatttgcagaaaataatgctcaaatg 540
Dbb 481 acaaaacaatcagccttttaaggtggctatcaagagatttgcagaaaataatgctcaaatg 540
Qy 541 atgaagtcagaaaaattttgttgaaactcacaagagagcacaataattatggcccgatataca 600
Dbb 541 atgaagtcagaaaaattttgttgaaactcacaagagagcacaataattatggcccgatataca 600
Qy 601 aatgagtatgaccagtagaattggaaattgtgtcctctgtgtaaaagcttatacaaaatgg 660
Dbb 601 aatgagtatgaccagtagaattggaaattgtgtcctctgtgtaaaagcttatacaaaatgg 660
Qy 661 gcagctcaaatggctgttaggtttgaaaaactgggttcccatggatcattgtgtaagcaagag 720
Dbb 661 gcagctcaaatggctgttaggtttgaaaaactgggttcccatggatcattgtgtaagcaagag 720

Dbb 661 gcagctcaaatggctgttaggtttgaaaaactgggttcccatggatcattgtgtaagcaagag 720
Qy 721 gatgctcctgatcctgtgatgatacttgtaagtcttactactcgaagggttccgctct 780
Dbb 721 gatgctcctgatcctgtgatgatacttgtaagtcttactactcgaagggttccgctct 780
Qy 781 aataagccttacaaacctaaatgtggacagaagtatggactggctgggtatcacgaattc 840
Dbb 781 aataagccttacaaacctaaatgtggacagaagtatggactggctgggtatcacgaattc 840
Qy 841 ggtggtccaattctccaagaccgacccgaagacattgattttcagttgacaggtttgtt 900
Dbb 841 ggtggtccaattctccaagaccgacccgaagacattgattttcagttgacaggtttgtt 900
Qy 901 cagaacaatggttcattctcaattactacatgatactatgaggaacaaaattttgcccgg 960
Dbb 901 cagaacaatggttcattctcaattactacatgatactatgaggaacaaaattttgcccgg 960
Qy 961 acatcatcagggttttctcattgcaactagctacgattatgatgctcctctcgatgaat 1020
Dbb 961 acatcatcagggttttctcattgcaactagctacgattatgatgctcctctcgatgaat 1020
Qy 1021 ggggttgctgaatgaaccaaaagtatgggcacttgagagacttacataaaggctatcaagcta 1080
Dbb 1021 ggggttgctgaatgaaccaaaagtatgggcacttgagagacttacataaaggctatcaagcta 1080
Qy 1081 tctgaaccggttttagtttctcatcatatgctgcggtgactagctcttgggaagtaatacagag 1140
Dbb 1081 tctgaaccggttttagtttctcatcatatgctgcggtgactagctcttgggaagtaatacagag 1140
Qy 1141 gctcatgtttatagatcaaaaactctggagcttgcgtgctgttttttccaaactatgacctct 1200
Dbb 1141 gctcatgtttatagatcaaaaactctggagcttgcgtgctgttttttccaaactatgacctct 1200
Qy 1201 agatattcaagtaaaagtcaactttcagaataaggccatacaactctgcctccatggctcactc 1260
Dbb 1201 agatattcaagtaaaagtcaactttcagaataaggccatacaactctgcctccatggctcactc 1260
Qy 1261 agcattctccgactcgaacaaactgcgttttacaacactgcacaggttaactctcctcaagc 1320
Dbb 1261 agcattctccgactcgaacaaactgcgttttacaacactgcacaggttaactctcctcaagc 1320
Qy 1321 tcgagctaaagatgacgcctgcaggtggtggattctcttggcagctacatacaatgaagaa 1380
Dbb 1321 tcgagctaaagatgacgcctgcaggtggtggattctcttggcagctacatacaatgaagaa 1380
Qy 1381 acgctactgctgatgacagcgatacaacttacagctaaacgactatgggaacagaaaac 1440
Dbb 1381 acgctactgctgatgacagcgatacaacttacagctaaacgactatgggaacagaaaac 1440
Qy 1441 gtcacaagagattcattcagactatctgtgtgacatgacaaaatgtaaatatagcatccta 1500
Dbb 1441 gtcacaagagattcattcagactatctgtgtgacatgacaaaatgtaaatatagcatccta 1500
Qy 1501 gaagatttctaaagacggaagatccttatctcactgttatgtccgctggctcactgctc 1560
Dbb 1501 gaagatttctaaagacggaagatccttatctcactgttatgtccgctggctcactgctc 1560
Qy 1561 ttgcatgttttcgtcaatggaaaactatcagaagactgtttatgtgataattggataatcca 1620
Dbb 1561 ttgcatgttttcgtcaatggaaaactatcagaagactgtttatgtgataattggataatcca 1620
Qy 1621 aaacttacatcagtgccacgctgaagttaagagctgtgtatcaaatgtctctgctc 1680
Dbb 1621 aaacttacatcagtgccacgctgaagttaagagctgtgtatcaaatgtctctgctc 1680
Qy 1681 agtgttccgttggtctcccgaaactgtggcgtgattatgatacatggaatgcaggagtt 1740
Dbb 1681 agtgttccgttggtctcccgaaactgtggcgtgattatgatacatggaatgcaggagtt 1740
Qy 1741 ctaggctcagctcagcttgagcgggtctcaatgaagggtcaagaacttggcgaacagagaa 1800
Dbb 1741 ctaggctcagctcagcttgagcgggtctcaatgaagggtcaagaacttggcgaacagagaa 1800


```
Qy 1801 tgggttaccaggttggtctgaagcgcaaatcggttaagtcttcaactccttaagtggagt 1860
Dy 1801 tgggttaccaggttggtctgaagcgcaaatcggttaagtcttcaactccttaagtggagt 1860
Qy 1861 tcttctgtgaatgggttcgaggttcaataatggtctcaaaagcagccctgaacttgggtac 1920
Dy 1861 tcttctgtgaatgggttcgaggttcaataatggtctcaaaagcagccctgaacttgggtac 1920
Qy 1921 aaggtctacattaacgcgccttggaggaaatgatcatcactagcttttagacatggcaagtatg 1980
Dy 1921 aaggtctacattaacgcgccttggaggaaatgatcatcactagcttttagacatggcaagtatg 1980
Qy 1981 ggaagagtcagatgatgataaaatgggtgaagcgttagtcgccattggcctggatcacata 2040
Dy 1981 ggaagagtcagatgatgataaaatgggtgaagcgttagtcgccattggcctggatcacata 2040
Qy 2041 gcacaaagcgactgcagcaaatgaagtattatgctggaacgttcaacagagaagaagtgcgga 2100
Dy 2041 gcacaaagcgactgcagcaaatgaagtattatgctggaacgttcaacagagaagaagtgcgga 2100
Qy 2101 actaactgcgacaacctctcagagatggtaccatgttccacagatcgtgctcaaaacca 2160
Dy 2101 actaactgcgacaacctctcagagatggtaccatgttccacagatcgtgctgcaaaacca 2160
Qy 2161 agtggaactttagtagtattcgaagaatgggaggtgaatcccaacaggaatttctcta 2220
Dy 2161 agtggaactttagtagtattcgaagaatgggaggtgaatcccaacaggaatttctcta 2220
Qy 2221 gtcaggagatcaagataaagaactgcgaagaagttaaaactgttcagtaactatggtgcttg 2280
Dy 2221 gtcaggagatcaagataaagaactgcgaagaagttaaaactgttcagtaactatggtgcttg 2280
Qy 2281 aattcgcgcgcaaaatcatatcacgaagctaaacatgaggtacagtttgcaaatgctc 2340
Dy 2281 aattcgcgcgcaaaatcatatcacgaagctaaacatgaggtacagtttgcaaatgctc 2340
Qy 2341 agctgaataaaacattagaagataaaagaataatttggataaaaggagtataaaatttac 2400
Dy 2341 agctgaataaaacattagaagataaaagaataatttggataaaaggagtataaaatttac 2400
Qy 2401 agagaatttcttattcttctgtaaaacttgggtttataaaagtttatcacagaatttctg 2460
Dy 2401 agagaatttcttattcttctgtaaaacttgggtttataaaagtttatcacagaatttctg 2460
Qy 2461 ttatttgattatgagattgagaagattgtacagcttccaaactactattagatcacaaa 2520
Dy 2461 ttatttgattatgagattgagaagattgtacagcttccaaactactattagatcacaaa 2520
Qy 2521 taaatttcatgtaaaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2554
Dy 2521 taaatttcatgtaaaaaaataaaataaaataaaataaaataaaataaaataaaataaaataaa 2554
```

RESULT 2

AAT01014
ID AAT01014 standard; cDNA; 2628 BP.

XX AC AAT01014;

XX XX

XX DT 10-FEB-1996 (first entry)

XX XX

DE Lupin exo-(1-4)-beta-D-galactanase cDNA sequence.

KW Exo-(1-4)-beta-D-galactanase; enzyme; transgenic plant;
crop improvement; ss.

XX Lupus angustifolius.

XX OS

XX Key Location/Qualifiers

FT CDS 130..2322

FT /*tag= a

XX

PN WO9523228-A1.
XX 31-AUG-1995.
XX 23-FEB-1995; 95WO-CB00372.
XX 23-FEB-1994; 94GB-0003423.
XX (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX Chengappa S, De SILVA J, Hellyer SA, Reid JSG;
PI WPI; 1995-311537/40.
DR P-PSDB; AAR82881.
XX Lupin and tomato exo-galactanase DNA and protein - useful for the
alteration of plant characteristic(s), e.g. texture, growth,
ripening
XX Claim 1; Page 25; 67pp; English.
XX This cDNA sequence may be used in the construction of a transgenic
plant in order to alter characteristics of e.g. growth, texture
or ripening of the plant or plant parts. This sequence may also be
expressed in a recombinant host for the production of the enzyme
which may be used for the modification, degradation or liquefaction
of plant materials in order to affect mechanical properties relating
to eating texture, particle sizes of, e.g. fruit or vegetable juices,
or extractability of colors, flavors or vitamins.
XX Sequence 2628 BP; 835 A; 448 C; 581 G; 764 T; 0 other;

Query_Match 47.2%; Score 1205.8; DB 16; Length 2628;
Best Local Similarity 73.5%; Pred. No. 5.7e-234;
Matches 1569; Conservative 0; Mismatches 557; Indels 9; Gaps 2;

Qy 98 ttgttttttggatttttttttcttcagtgaaagctagtttcttctatgatgacagacta 157
Dy 194 tgtattgttttttgggtttgtatgtcacagctctgttacttatgatcataaagcca 253
Qy 158 taatcataatgggaaaaaattcttattcttggttcaattctattatccaagaagca 217
Dy 254 ttatgattaaaggcagagaagaatttgcctctgttccattctactatccaagaagca 313
Qy 218 ctccacagatgtggcctgatcttatacaaaagctcaagatggaggttagatgtattg 277
Dy 314 cactcagatgtggccagaccttattcaaaagccaaagatggaggttgatgtattag 373
Qy 278 aaacttatgtttctggatggacatgagccttctccttgaaaaataataatttgaagaa 337
Dy 374 agacttatgttctggaaatggacatgacaccttctccttgaaaaataattatttggagata 433
Qy 338 gatatgatctttagattcatcaaaatgggtacaaagcagaggaactttatgtcaatttac 397
Dy 434 ggttgacctgtgtgttcataaaagtgttcacagacgtgtctatttgtctatcctca 493
Qy 398 gtattggcccttaactgtgtgctgaatggaaacttgggggattccctgtttggctaaaaat 457
Dy 494 ggaattgctcttcatatgtgctgaatggaaacttggaggatttctctgttggcacaat 553
Qy 458 atgtgcttggtatgaatttagaacaacaacatcagcctttaaagtggtggtcgaagat 517
Dy 554 atgttcttggtatttcttccagaacacacatgagccttccaggaggaatgcacaaat 613
Qy 518 ttgttcagaaaaatagtcacatgatgaatgcagaaaaatttttgaattctcaaggagac 577
Dy 614 tcactgagaagattgtataataatgaagcagagaagtgttctcaatccccaggaggtc 673
Qy 578 caataattatggccagatatacaaaatgatgtgaccagtagaattggaaaattggtgctc 637
Dy 674 caataattctgtctcagatagagaatagtagtgaccagtagaattggaaaattggtgctc 733

Db 904 tcaattattacatgtatcatgtgaggaaacaaactttggacgactgtctggtgtccattta 963
Qy 980 ttgaaactagctaacgatttatgtctctcgtcgtgaataatggttgcgtgaatgaacaa 1039
Db 964 ttgctactagtattgactatgacacacacttgatgaataatggtattatgagaaacaa 1023
Qy 1040 agtatggcaacttgagagacttacataaagaattatcaaatctatctatgaaacggttttagttt 1099
Db 1024 aatgggttcacctgaagatctgcatagagcaataaagctttgtgaaccagcttttagtct 1083
Qy 1100 catcatatgctgcggtgactagcttggaaagtaatacaagagcgctcatatttatagataca 1159
Db 1084 ctggagatccagcgtgacagcacttggaaccagcagagggcccatcttttttaggtgga 1143
Qy 1160 aatctggagcttggctgcttttttccaaactagctctagatattcattcagtaaaagtcga 1219
Db 1144 aggctggtcttgtgctgctattcctgttaactacgacccaactcttttgcactgctgt 1203
Qy 1220 cctttcagaataggccatacaatctgctccatggtccatcagcaattcttccgactgca 1279
Db 1204 catttgcaaacaggcattacacacttgccaccatggtccaatcagcattcttccgactgca 1263
Qy 1280 aaactgcoctttacaaactgcacaggttaactctcaagctcgagcaataagatgaagc 1339
Db 1264 agaacactgtatttaatacacagcaggtcgtgctcaagtgctcagatgaagatgactc 1323
Qy 1340 ctgcaggtggtgattctcttggcagtcatacaatgaagaaacgcctactgctgatgaca 1399
Db 1324 cagtcagcagaggttgccttggcagtcattcaatgaagagacatcatcttatgaagaca 1383
Qy 1400 gcgatacacttacagctaacgagctatgggacagaaaaacgctcacaaagagattccatcag 1459
Db 1384 g---tagttttacagtttgggtctattggaacagataaatacaacaagagacgtgtctg 1440
Qy 1460 actatctgtgtacatgacaataatgaataatagcatctcaatgaaggtatttctaaagaagc 1519
Db 1441 attatttgggtattcaacagatgccaagattgattcgaagagaaaagtttttgagagcg 1500
Qy 1520 gaaaggaaccttatctcactgttatgtccgctggtcattgtctgttcttctcgaatg 1579
Db 1501 gaaaatggccttggcttacgcatgtcagctggcagtcgcatgttcattgttttggatg 1560
Qy 1580 gaaactatcaggaactgtttatggtacattggtatgaataatccaaaacttacatcacagtgga 1639
Db 1561 gtcaattagcgaagactgcatggaagtttagaaaaaccgaaactaactttcagtaaaag 1620
Qy 1640 acgtgaagttaagagctggtattacaagatttctcgtcaggtttccgttggctctcc 1699
Db 1621 ccgtaaactctgagagcaggtgttaacaagatttctcactgagcattgctgttggccttc 1680
Qy 1700 cgaacttggcgtgcatcatgatacatggaatgagcaggttcttagtccagtcacgttga 1759
Db 1681 cgaatatcgcccaattttgagacatggaatgctggtgttcttggccagctcactaa 1740
Qy 1760 gcggtctcaatgaaggggtcagaacttggcgaaacagagaataatggtcttacaaggttggtc 1819
Db 1741 ctggctgtgacgaggggaaagagatttaacatggtgcagaaatggttctacaaggttggtc 1800
Qy 1820 tgaagcgcaactgttaagttactcactcttaagtggagttcttctgttgaatgggttc 1879
Db 1801 taaaaggagacccctgagttcttcaactcactcagttgtagcccatccgtggagtggttg 1860
Qy 1880 gaggttcaatggtctcaagcagccctgacttggtaacaggtcactacatttaacgcgc 1939
Db 1861 aagctcttttagtgacagagcagcgaactcagttggtataagactacattcaatgctc 1920
Qy 1940 ctggaggaaatgatccactagcttttagacatggcaagttatgggaaaaaggtcagatagga 1999
Db 1921 cagatggaaatgaacctttgcttttagatgaataccatgggcaaaaggtcaagtatgga 1980
Qy 2000 taaatggtgaagcgtaggtgcgaattggcattggtgatacatagcacaaagcgactgcagca 2059
Db 1981 taaatggtcagagcctcgagccactggcctgcatataataatcatctcggaagttgagt 2040

Qy 2060 aatgcagttatgtggaacttcaacgagagaagatgcggaactaactcgcgacaacctt 2119
Db 2041 tctgtataactactgctggtttgatgagaaaaagtgcctaaactaactgtggtgaggct 2100
Qy 2120 ctccagagatggtacacatgttccacgactgctggtgaaacccaagtggaaacttgttagtag 2179
Db 2101 cacaagaatggtaccacgtaccocggtcttgggtatctactactggaattttagttg 2160
Qy 2180 taticgaagaatgggaggttaatccaacagagaatttctctagtcagagatcaagataaa 2239
Db 2161 tattcaggaatgggagagatcccttatggaatcacttttagtcaaaagagaatagga 2220
Qy 2240 gaactcgaataaagtaaaacttgttccagtaactatggtctgaattcgc 2287
Db 2221 gtgtttgtgcgatataatgatgggcaaccacagtattgaattggc 2268

RESULT 5

AAQ89602
ID AAQ89602 standard; cDNA; 2945 BP.

XX AAQ89602;

XX 21-NOV-1995 (first entry)

XX Tomato beta galactanase coding sequence.

XX Beta galactanase; galactosidase; fruit; ripening; tomato;
firmness; texture; viscosity; genetic engineering; ss.

XX Lycopersicon esculentum.

PN W09510622-A.

XX 20-APR-1995.

XX 10-OCT-1994; 94WO-GB02203.

XX 12-OCT-1993; 93GB-0020930.

XX (ZENE) ZENECA LTD.

XX Bird CR, Carey AT, Holt KA, Picard S, Schuch WW;

PI Seymour GB, Tucker GA;

XX WPI; 1995-161807/21.

DR Modifying cell wall metabolism in plants with a DNA construct -
that affects galactosidase activity, esp. for control of fruit
ripening, also transformed plant cells and derived plants

PS Claim 3; Page 30-31; 39pp; English.

XX Cell wall metabolism in plants can be modified by transforming the
plants with a DNA construct encoding a tomato galactanase which is
able to modify the activity of at least one galactosidase. Plants
transformed in this way have latered galactanase gene expression
CC which alters fruit-ripening characteristics. Reduction of gene
expression will increase fruit firmness, resistance to mechanical
damage, improve texture and resistance to disease and also increase
CC the viscosity of processed fruit. Increasing galactosidase activity
CC will also modify fruit texture and processing properties. This is
CC the tomato galactanase enzyme coding sequence.

XX Sequence 2945 BP; 851 A; 525 C; 699 G; 870 T; 0 other;

Query Match 45.7%; Score 1168.2; DB 16; Length 2945;

Best Local Similarity 70.3%;

Matches 1579; Conservative 0; Mismatches 663; Indels 3; Gaps 1;

Qy 43 tttttttgaaatgtagaaaaatgctaaggagactaatgtgtgtgtattattgtt 102

Qy 2263 cagtaactatgtgtctgaattcgc 2287
 Db 2268 tggcaaccacagtattgaattgc 2292

RESULT 6

AAZ29338
 ID AAZ29338 standard; cDNA: 3224 BP.

AC AAZ29338;
 XX

DT 29-FEB-2000 (first entry)
 XX

DE Tomato beta galactosidase-1 cDNA.
 XX

KW Tomato beta galactosidase-1; TBG-1; clone pZBG2-1-4; fruit ripening;
 KW Rutgers tomato plant; pectin; biofilm; gene mapping; ds.
 XX

OS Lycopersicon esculentum.
 XX

XX Key Location/Qualifiers
 FH 306..2813
 FT CDS /*tag= a

FT sig_peptide /product= "Tomato beta galactosidase-1"
 FT 306..377

FT mat_peptide /*tag= b
 FT 378..2810

FT /*tag= c
 XX

PN W09964564-A1.
 XX

XX 16-DEC-1999.
 PD

XX 08-JUN-1999; 99WO-US12697.
 PF

XX 09-JUN-1998; 98US-0088805.
 PR

XX (USDA) US DEPT OF AGRICULTURE.
 XX

XX Gross KC, Smith DL;
 PI

XX WPI; 2000-097532/08.
 XX

DR P-PSDB; AAY44303.
 DR

XX New beta-galactosidases, used to prepare transgenic plants with altered
 PT fruit ripening -
 PT

XX Claim 2; Fig 2; 85pp; English.
 PS

XX The present sequence encodes tomato beta galactosidase-1 (TBG-1). This
 CC is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from
 CC breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants.
 CC TBG cDNA has been expressed in E.coli strain XLblue MR. This is used
 CC for modifying cell wall metabolism and controlling ripening of fruit by
 CC altering activity of beta galactosidase II protein. Pectin with reduced
 CC galactosyl content is produced for use in biofilms or solutions.
 CC Fragments of the cDNA are used for gene mapping and for detecting
 CC expression of beta galactosidase in plant tissue. Transgenic plants
 CC with altered fruit ripening are produced by introducing DNA constructs
 CC comprising a TBG cDNA associated with promoter/enhancer elements.
 XX

SQ Sequence 3224 BP; 981 A; 556 C; 738 G; 949 T; 0 other;

Query Match 45.7%; Score 1168.2; DB 21; Length 3224;
 Best Local Similarity 70.3%; Pred. No. 2.3e-226;
 Matches 1579; Conservative 0; Mismatches 663; Indels 3; Gaps 1;

Qy 43 ttttttttgatgtggaaaaaatgctaaggactaatgtgtgtgtgtattattgtt 102
 Db 282 ttgattcttgattgttgacaaaatgggttttggatggcaatgtgtgtgtgtt 341

Qy 103 ttattggatttttttcttcagtgaaagctagtggtttcttatgatgacagagctataac 162
 Db 342 ttgtgttatgggtttcttggaattgctctctctgtttcatatgaccataaaagctatcat 401
 Qy 163 ataaatgggaaaaaataattcttattcttggttcaattcattatccaagaagcactcca 222
 Db 402 gtaaatggacaaaagaaaattctctattctggtatccattcactaccctagaagcaccct 461
 Qy 223 cagatgtgctgctgattctatacaaaaagctaaagagtgaggttagattttatttataaact 282
 Db 462 gagatgtggtccagatccttattcagaagaagaaaagggagtggttatacagact 521
 Qy 283 tatgtttctggaatggacatgagccttctctctggaataataattttgaaggaagatat 342
 Db 522 tatgtttctggaatggcagtgagcctgaagaagggaaaataattattttgaagagaggtat 581
 Qy 343 gatctgttagattcacaataatgggtacaaagagcagacttttatgtaaatctacgtatt 402
 Db 582 gatttagtaagtctataaagtgtgcaagaagcagacttttatgtaaatctacgtatt 641
 Qy 403 ggcctttagctctgtgctgaatggaactttgggggattccctgttttgcttaaaatatgtg 462
 Db 642 ggaacttatgcatgctgctgaatggaaattttgggggtttctctgtctggtcgaagcatgt 701
 Qy 463 cctggtatggaatttagaacaacaatcagccttttaagggtgctatcgaaggtattgtt 522
 Db 702 ccaggtatttagttcagaacaacaatgagccattcaaggctgcaatgcacaaagtctact 761
 Qy 523 cagaaaatagtoaacatgatgaagtgcagaaaatttggttgaatctcaaggaggacacata 582
 Db 762 actaagattgttgatgatgaaagcagaaaagctctatgaaactcagggttggtccaatt 821
 Qy 583 attatggccagatacaaaatgagtatgaccagtagtaatgggaattgggtcctcgtgt 642
 Db 822 attctatctcagatagaaatgaatggactatggagcttggagtggaactggagaaactgtg 881
 Qy 643 aaagcttatacaaaatgggcagctcaaatggctgagttttgaaactgggtgccctcagg 702
 Db 882 aaagtttactcagaatgggcagcacaatggctggtggtatctggcactgggtgccctcagg 941
 Qy 703 atcatgtgaagcaagagatgctcctgctgctgctgctgctgctgctgctgctgctgctac 762
 Db 942 atcatgtgaagcaagatgctcctgctgctgctgctgctgctgctgctgctgctgctac 1001
 Qy 763 tgcgaaggttcgctccttaataagccttacaaactcaaaactgagtgagcaagatagact 822
 Db 1002 tgtgactacttcacacaaaataaaggcttaataacccagatgtgagactggaagcctggaca 1061
 Qy 823 ggctgggtatcacgaattcgggtgtccaatctctcaaaagaccagccgaagacattgcattt 882
 Db 1062 gccgtgtttaccgaatttggaggtccagttccttaccgtcctgcagagatagtcattt 1121
 Qy 883 tcagttgccaggtttgttcagaaacaatgggttcattctctcaattactacatgataatgga 942
 Db 1122 gctgtcgaagatttatcaaacggaggtcctctcattcaattactatgataatgataatgga 1181
 Qy 943 ggaacaaattttggccgacatcatcagggttttcattgcaactgacatgattatcat 1002
 Db 1182 ggaacaaactttggaagacttctgggtggccatttatgtctactagttagattatgat 1241
 Qy 1003 gctcctctcgatgaatatgggtgtgctgaatgaaccaaagtatgggcacttgagagactta 1062
 Db 1242 gcaccttagatgaatttgggtcattacggcagcctaataatgggtcatctgaaagactta 1301
 Qy 1063 cataaagctataagctatctgaaccggccttttagtttcatcatatgctgcggtgactagt 1122
 Db 1302 catagagcaataaagctctgtgagccagcttttagtctgtagatcccaactgtgacatcc 1361
 Qy 1123 cttgggaagtaatacgaagctcatgttttatagatacaaaatctggagctgtgctgctttt 1182
 Db 1362 ttadgaaactatcagaagacgtgttttcaagtacagctgtgggtgcgtgcgtcttc 1421
 Qy 1183 ttatccaaactatgactctagatattcagtaaaagtcaccttttcagaatagcgccatacaat 1242

D	b	1422	ctagcaaatcacaccagcactctttgctaagtgccatttggaacatgcattataac	1481
Q	y	1243	ctgctccatggctccatcagcattctccgactgcacaaactgcgtttacaacactgca	1302
D	b	1482	ttgccacctggctctatcagcattctccgactgcagaacacactgctataactgca	1541
Q	y	1303	caggttaaactctcaagctcgagcataaagatgacgctgcaggtgggtgattcttgg	1362
D	b	1542	agggttgggtctcaagtgctcagatgaagatgactccagtcagtagagattctcatg	1601
Q	y	1363	cagtcatacaatgaagaacgctactgctgtagcacgcatatacaactacaagtaacgga	1422
D	b	1602	gagtcattcaatgaagacgagcatgcgatgaaga--cgacactttcacagttgttggg	1658
Q	y	1423	ctatggacacagaaaaacgtccaagagattcatcagactatctgtgtacatgacaaat	1482
D	b	1659	ttatcgagcagattaataccaaagagatgatctgtattacttgttggtatagactgac	1718
Q	y	1483	gtaaatagcatctaatgaagattctcaagaacggaagggatccttatctcaactgtt	1542
D	b	1719	atggagattgatccaaacagaagattttgaatagtgaattggccttggttaactgtc	1778
Q	y	1543	atgtccgtgggtcatgcttggtaattttctgttcaatggaaacatcatcaggaaactgtt	1602
D	b	1779	ttttctgctggcatgcattgcatgtattctgtgaatggttcaatlagcaggaaactgtac	1838
Q	y	1603	ggtaacttgataatccaaacttcatacagtgccaagtgcgaactgaagttaagagctggt	1662
D	b	1839	ggaggtctagaacacccaactcaacttcagcaacggtataaatctcgagagctggctg	1898
Q	y	1663	aacaagattctctgctcagtgctttccgttggtctccgaacgttggcgtgcattatgat	1722
D	b	1899	aacaagattctctgaagcattgctgtgtgtcttcgaacggtggccctcatttggag	1958
Q	y	1723	acatggaatcgaggaggtcttaggtcgaatgcagttgagcgttgcattgcgaatgaaggtc	1782
D	b	1959	acatggaatgctggtcttcttgaccagtttcacctaattggacttaataaggagcaacga	2018
Q	y	1783	aacttggcgaacagaaaatggctctacaaggttgcctgaagcgcaatcgttaagcttt	1842
D	b	2019	gatttaacatcggcagaaaaggcttctacaaggctggctcaaaagagagccctgagctt	2078
Q	y	1843	cactccttaagtggaggttctctgttgtaatgggttcaggttcacttaatgggtcacaag	1902
D	b	2079	cattcactcagtgtagcccatccgtggagtggtgggaagcgtctttagtggtcagaag	2138
Q	y	1903	cagccctgacttggtcagaaggtacatttaacgcgcctggaggaaatgatccactagct	1962
D	b	2139	cagccactcagtggtataagactacattccaatgctccagatggaatgaaccttggct	2198
Q	y	1963	ttagacatggcaagtatgggaaaaagtcagatatggataaatgtggaagcgtaggtcgc	2022
D	b	2199	ttagatatgaataccatgggcaaggtcaagtatggataaatgtgtcagagcctcgagcgc	2258
Q	y	2023	cattggcctggatcacatagacacaagcgactgcacaaatgcagttatgctggaacgttc	2082
D	b	2259	cactggcctgcatataaaactctggaagtgtagtctgttaactatacttggctgggtt	2318
Q	y	2083	aacgagaagaatgctgggaactaactcgcggacaaacctctcagagatggtaccatgtcca	2142
D	b	2319	gatgagaaaaagtcctaactaaccttggtgaggtgcacaaagatggtaccacgtacc	2378
Q	y	2143	cgatcgtggctgaaacacagtgtaaacttggtagtagtatctcgaagaatggggaggtta	2202
D	b	2379	cggtcttggctgtatcctactggaatttggtagttgtatttcgaggaatggggaggtat	2438
Q	y	2203	ccaacaggaattctctagtcaggagatcaagataaagaacctcgaagaataaactgtt	2262
D	b	2439	cccttaggaatacactttagtccaaagagaaaataggggaggtcttctgctgatatacg	2498
Q	y	2263	cagtaactatgggtgcttgaaattcgc	2287

Db	2499	tggaaccacagttattgaattggc	2523
RESULT	7		
AAC49769			
ID	AAC49769	standard; DNA; 2469 BP.	
XX	AC		
XX	AAC49769;		
XX			
DT	18-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana	DNA fragment SEQ ID NO: 62367.	
XX			
KW	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway;		
KW	metabolic pathway; promoter; termination sequence; ss.		
XX			
OS	Arabidopsis thaliana.		
PN	EP1033405-A2.		
XX			
PD	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-01231825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	16-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139492.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	

[illegible]

Db	1226	cagaggtttaacaactgc	aaaggctcaatg	cccaagcg	tacacaggaata	tggttcta	1282			
Qy	1343	caggtggtgattctct	ctggcagtcataca	atgaagaacgc	tactgctgatgacagcg	1402				
Db	1286	ctggtactaaattctct	ctgggacatcat	tcaacgaagcgact	ccctacagcgaatgaagctg	1345				
Qy	1403	ataccttacagctaac	gcgactatgg	agacgaaacacg	tccacaagagattcatcagact	1462				
Db	1346	gtacgtgtcgaagaa	cggctgttg	gaacaaataag	catgacttgggacaaatctgact	1405				
Qy	1463	atctgtgtacatga	caaaatgta	atataagcat	ctaaatgaaggattctcaagaacggaa	1522				
Db	1406	atcttgggtataa	acagacattaca	cgttccog	tggagacatttttgaagactggtg	1465				
Qy	1523	aggatcttatctca	ctgttatgtc	gcgtggtcatg	ctctgcattgttttcgtaattgaa	1582				
Db	1466	attcacctctctta	cagttatgtc	agctggacatg	ctctctcatgtgtttgttcaatggtc	1525				
Qy	1583	aactatcaggaact	gtttatgtg	tacattggata	atccaaaactttacatacagttgccaag	1642				
Db	1526	agcttcaggaa	ctgttatg	gagacttg	accaccaactaaactttagccagaaga	1585				
Qy	1643	tgaagttaagagc	gtgtataa	caagattctct	ctgcagtggttcogttggtctccoga	1702				
Db	1586	tcaactacatcag	gtgtcaacag	attgctctct	gagtggttcgactgggtctccoga	1645				
Qy	1703	acgttggcgtgc	attatgat	atcatggaat	gcagagttcttaggtccagtcacgttgagcg	1762				
Db	1646	acgttggacac	attttgaac	agtggaataa	aggaggttcttctgttcoggtccacttgaagg	1705				
Qy	1763	gtctcaatgaag	gttcaaga	aaacttggcg	aaacagaaatggtcttcaaaagttggtctga	1822				
Db	1706	gagtaaacac	gagaacat	gggacatgt	cgaaatgaaatg	gtcctataagattggtgtaa	1765			
Qy	1823	aaggcgaatc	ogttaagct	ctcactcctta	agtgggaggttctctgttgaatgggttcogag	1882				
Db	1766	aggtgaggtt	tgcttgc	atccacacag	agaggtccoggtgtgagatggactcaag	1825				
Qy	1883	gttcaactaat	ggtc	aaagacgccc	tgacttgg	tacaaaggtacatttaacgcgcctg	1942			
Db	1826	gatcatttgg	cggaagag	caaccattaa	oc	gtgtacaa	gtcactgcttgcacaccag	1885		
Qy	1943	gaggaaatgat	ccactag	ctttag	acatggc	aatgtggg	aaaaaggttcagatatggataa	2002		
Db	1886	caggaaatga	accatttgg	ctttag	acatg	aacacaat	gggaaaaaggttcaagtttggataa	1945		
Qy	2003	atggtgaaggc	gtaggtcg	ccattgg	ctcgtgata	catagca	aggcgagctgcagcaaat	2062		
Db	1946	acgttaga	aatattg	gacgttcatt	gtgcttg	catataa	agctccaagggtatcttgtggacgtt	2005		
Qy	2063	gcagttatg	ctggaa	cgttcaac	gagagaag	tgggaact	aaactaac	ctcgggacaaccttctc	2122	
Db	2006	gtac	tacgtgtga	acttcc	gacgaaga	aatgctta	ag	tatgttgg	tgaagcttccac	2065
Qy	2123	agagatggt	taccaat	gttcca	gatcgttgg	ctgga	aaccaag	tggaaaactt	gtgtagtagtat	2182
Db	2066	aagatggt	taccaat	gttcca	gatcgttgg	ctgga	aaccaag	tggaaaactt	gtgtagtagtat	2122
Qy	2183	tcgaagaa	tggggagg	taata	cccaac	gagaaat	ttctct	tagtcagg	agatcaagataa	2238
Db	2123	tcgaagaa	tgttag	gtgtg	atctca	acgaaat	ctctt	gtgtta	aaagacataaaaa	2178

```

RESULT      8
AAC49810
ID   AAC49810 standard; DNA; 2571 BP.
XX
XX
AC   AAC49810;
XX
DT   18-OCT-2000 (first entry)
XX
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 62521.
XX
XX

```


Qy	1520	gaaaggatccttatotcactgttatgtccgctgtgtcatgtcttgcattgttttcgtcaatg	1579	XX	25-FEB-2000;	2000EP-0301439.
Db	1484	gagagctaccaactctcattatcaatctacagccatgcgtgcataatttgcattg	1543	XX	25-FEB-1999;	99US-0121825.
Qy	1580	gaaactactcaggaaactgtttatgtgtacattggaataccaaacttaccatcagtgaca	1639	PR	05-MAR-1999;	99US-0123180.
Db	1544	gacagcttcagggttctgcttggaaacaggcaaaacaggagattttactatcaggga	1603	PR	09-MAR-1999;	99US-0123548.
Qy	1640	acgtgaagttaagagctggtattacaagatttctctgctcagtggttccgttgcctcc	1699	PR	23-MAR-1999;	99US-0125788.
Db	1604	agatacatctgcattctcggaaactacaagaattcgtgtcgtgagtggtcttggatcac	1663	PR	29-MAR-1999;	99US-0126264.
Qy	1700	cgaactgtggcgtgcattatgatacatggaatcagaggagttctcaggtccagtcacgttga	1759	PR	01-APR-1999;	99US-0126785.
Db	1664	caaatgtgggtggacactttgagtcagtgaatactggaatcttgggtccgtggtcattgc	1723	PR	06-APR-1999;	99US-0127462.
Qy	1760	gcgtctcaatgaagggttcagaaacttggcgaacagaaatgtcttacaagggtgtgc	1819	PR	08-APR-1999;	99US-0128234.
Db	1724	atgggttgagtcagggaagaaatgatttgcattgcatggaataatggaataccaggtggac	1783	PR	16-APR-1999;	99US-0128714.
Qy	1820	tgaaggcgcaatcgttaagtcttcaactcttaagtggagttcttctgtgaatgggttc	1879	PR	19-APR-1999;	99US-0129845.
Db	1784	tgaagggtgaagctatgaattctgcatttccaacacactccctctctatttggatggatgg	1843	PR	21-APR-1999;	99US-0130077.
Qy	1880	gaggttcactaatggctcaaaag---cagccctgacttggatgatacgaaggtacatttaacg	1936	PR	23-APR-1999;	99US-0130510.
Db	1844	atgcgtcttaactgtacaaagcctcagccttggacatgacaaagacttatttgcattg	1903	PR	28-APR-1999;	99US-0130891.
Qy	1937	cgctggaggaaatgatccactagcttttagacatggcgaagtatgggaaggtcagatat	1996	PR	30-APR-1999;	99US-0131449.
Db	1904	cgccgaaggaaatgagcgccttcttggacatggaaggatgggaaggtcagattt	1963	PR	04-MAY-1999;	99US-0132048.
Qy	1997	ggataaatggtgaaggcgtaggtcgccattgcttgatcatatagcacaaagcgaactga	2056	PR	05-MAY-1999;	99US-0132407.
Db	1964	gggtgaacggtgagagcatcggagatactggacagcat---tgcgactggtgactgca	2020	PR	06-MAY-1999;	99US-0132485.
Qy	2057	gcaaatcgatttctgtggaacgttcaacgagaagaagtcggaactaactgcggacac	2116	PR	07-MAY-1999;	99US-0132486.
Db	2021	gtcattgtagctacacgtggaacataataaaccgaacaaatgccaaaccgctgtggtcaac	2080	PR	11-MAY-1999;	99US-0132863.
Qy	2117	cttctcagagatgttaccatgttccacgactggtggctgaaaccgaagtggaactgttag	2176	PR	14-MAY-1999;	99US-0134256.
Db	2081	ctacacaaaggtgtaccatgtcgcgagagctgtgttaaagccgagcacaatctgttgg	2140	PR	18-MAY-1999;	99US-0134218.
Qy	2177	tagtatcgaagaatgggaggtgaatccaacaggaatttcttagtcaggagatcaaat	2236	PR	14-MAY-1999;	99US-0134219.
Db	2141	tcatctcagaggagcttgggtggaacccatcaactgtctcttggatcaaatcagcat	2200	PR	14-MAY-1999;	99US-0134221.
Qy	2237	aaagaactcg 2246		PR	14-MAY-1999;	99US-0134370.
Db	2201	caggagtttg 2210		PR	18-MAY-1999;	99US-0134768.
RESULT	9			PR	19-MAY-1999;	99US-0134941.
AAC46270				PR	20-MAY-1999;	99US-0135124.
ID	AAC46270	standard; DNA; 2553 BP.		PR	21-MAY-1999;	99US-0135353.
XX				PR	24-MAY-1999;	99US-0135629.
AC	AAC46270;			PR	25-MAY-1999;	99US-0136021.
XX				PR	27-MAY-1999;	99US-0136392.
DT	18-OCT-2000	(first entry)		PR	28-MAY-1999;	99US-0136782.
XX	Arabidopsis thaliana	DNA fragment SEQ ID NO: 49529.		PR	01-JUN-1999;	99US-0137222.
DE				PR	03-JUN-1999;	99US-0137528.
XX				PR	04-JUN-1999;	99US-0137502.
KW	Hybridisation assay; genetic mapping; gene expression control;			PR	07-JUN-1999;	99US-0137724.
KW	protein identification; signal transduction pathway;			PR	08-JUN-1999;	99US-0138094.
KW	metabolic pathway; promoter; termination sequence; ss.			PR	10-JUN-1999;	99US-0138540.
OS	Arabidopsis thaliana.			PR	10-JUN-1999;	99US-0138847.
XX	EPI033405-A2.			PR	14-JUN-1999;	99US-0139119.
PN				PR	16-JUN-1999;	99US-0139452.
XX	06-SEP-2000.			PR	16-JUN-1999;	99US-0139453.
PD				PR	17-JUN-1999;	99US-0139492.
				PR	18-JUN-1999;	99US-0139454.
				PR	18-JUN-1999;	99US-0139455.
				PR	18-JUN-1999;	99US-0139456.
				PR	18-JUN-1999;	99US-0139457.
				PR	18-JUN-1999;	99US-0139458.
				PR	18-JUN-1999;	99US-0139459.
				PR	18-JUN-1999;	99US-0139460.
				PR	18-JUN-1999;	99US-0139461.
				PR	18-JUN-1999;	99US-0139462.
				PR	18-JUN-1999;	99US-0139463.
				PR	18-JUN-1999;	99US-0139750.
				PR	18-JUN-1999;	99US-0139763.
				PR	21-JUN-1999;	99US-0139817.
				PR	22-JUN-1999;	99US-0139899.
				PR	23-JUN-1999;	99US-0140353.
				PR	23-JUN-1999;	99US-0140354.
				PR	24-JUN-1999;	99US-0140695.
				PR	28-JUN-1999;	99US-0140823.
				PR	29-JUN-1999;	99US-0140991.
				PR	30-JUN-1999;	99US-0141287.
				PR	01-JUL-1999;	99US-0141842.
				PR	01-JUL-1999;	99US-0142154.
				PR	02-JUL-1999;	99US-0142055.
				PR	06-JUL-1999;	99US-0142390.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

Db 1685 tcttgggtccggtggcattgcatgggttagtcaagggaataatggatttgcattgacaga 1744
QY 1799 aatggcttaccagggttgcctgaagcgcaatcggttaagtcttcaactcttaagtggga 1858
Pt 1745 aatggacataccagggtggagactgaaggtgaagctatgaatttgcatttccaacaca 1804
XX
QY 1859 gtctctctgtgaatgggtgcaggttcactaatggtccaaaag---cagccctgactt 1915
Db 1805 ctccctctattgtagtgatggatgcgtccttaactgtacaaagcctcagccttgaact 1864
QY 1916 ggtacaagctacatttaacgcgcctggaggaataatgataccactagctttagacatggcaa 1975
Db 1865 ggcacaagacttattttagtcgcccgaaggaataatgagcgttcttggcatggaag 1924
QY 1976 gtagtggaaaaggtcagatcatgataaattggtgaagcgtaggtgcctcattggcctggat 2035
Db 1925 gaatgggaaaaggtcagatttgggtgaacggtgagagcatcggaagatactggacagcat 1984
QY 2036 acatagcacaaagcgactgcagcaaatgcagttatgctggaacgttcaacgagaagaat 2095
Db 1985 ---ttgcgactggtgactgcagtcattgtgctacactggaacatataaacgacaat 2041
QY 2096 gcggaactaactgcgacaaacttctcagagatggtaccatgttccacagatcgtgctga 2155
Db 2042 gccaaaccgctgtgtcaacctcacaaaaggtgtaccatgtgcgagagcttggttaa 2101
QY 2156 aaccaagtggaaactgttagtagtattctcgaagaatgggaggttaattccaacaggaattt 2215
Db 2102 agccgagcaaaactgtgtgctcatttcgagagcgttggtagaaccccatcaactgtct 2161
QY 2216 ctctagtcaggagatcaagaataaagaaactcg 2246
Db 2162 ctttggtaaaaagatcagtatcaggagtttg 2192

RESULT 11

AAT15173
ID AAT15173 standard; cDNA; 2778 BP.
AC AAT15173;
XX
XX
DT 02-JUL-1996 (first entry)
XX
DE Nasturtium xyloglucan-specific beta-galactosidase cDNA.
XX
KW Xyloglucan-specific beta-galactosidase; nasturtium; gelling;
KW transgenic plant; ripening; ss.
XX

OS Tropaeolum majus.

XX Key Location/Qualifiers
FT CDS 23..2596
FT /*tag= a
FT sig_peptide 23..121
FT /*tag= b
FT mat_peptide 122..2593
FT /*tag= c

W09607743-A1.

XX
XX
PD 14-MAR-1996.
XXXX 07-SEP-1995; 95WO-GB02098.
XXXX 09-SEP-1994; 94EP-0306635.
XXXX (UNIL) UNILEVER NV.
XXXX (UNIL) UNILEVER PLC.
XXXX Chengappa S, De Silva J, Hellyer SA, Reid JSG;
XXXX WPI; 1996-171617/17.
XX

DR P-PSDB; AAR95287.
XX
XX New isolated xylo-glucan-specific beta-galactosidase and related DNA
PT - used for altering the gelling properties of xylo-glucan or for
PT altering plant characteristics e.g ripening
XX
PS Claim 7; Fig 1; 36pp; English.
XX
CC A cDNA clone (AAT15173) codes for a nasturtium xyloglucan-specific
CC beta-galactosidase (XSBG) (see AAR95287) that catalyses the
CC hydrolysis of terminal galactose residues from polymeric xyloglucan.
CC It was obt. from a cDNA library of germinating nasturtium seeds
CC using primers (see AAT15175-82) designed on the basis of isolated
CC peptides of XSBG. The cDNA can be incorporated into a vector
CC for use in recombinant prodn. of mature XSBG, or can be used to
CC alter plant characteristics, e.g. growth, texture and ripening.
XX
SQ Sequence 2778 BP; 820 A; 519 C; 652 G; 787 T; 0 other;

Query Match 27.8%; Score 708.8; DB 17; Length 2778;
Best Local Similarity 59.9%; Pred. No. 1.le-133;
Matches 1297; Conservative 0; Mismatches 822; Indels 45; Gaps 5;

QY 80 tdtgtgttattagttattgtttattgttgatttttttcttcagtgaaagtagtgttt 139
Db 75 ttttgcgttctgttctgaattgttgggtgtgtttttttcttg---catcacaatgta 131
QY 140 cttatgatgacagactaataataatgggaagaaaataattttattttctgttcaa 199
Db 132 cttatgatcgtcgtctctcatcattgtggcgaagggaggtcattgattcagcttcta 191
QY 200 ttcatattccaagaagcactccacagatgtgcctgatcttatacaaaaggctaaagatg 259
Db 192 ttcataccctcgaagtgttctcgtgaatgtgcccggggttttccaacagctaaagaag 251
QY 260 gagggttagatgttattgaaacttatgttttcttggaaatgacatgagccttctctggaa 319
Db 252 ggggtatgcagctgacgaatcttatgttcttggaaatgtcatgagctctccaggaa 311
QY 320 aataaatttgaaggaagatgacttcttgatattcattcaataatggtacaaagacag 379
Db 312 aataaatttcgaggggcgatcagctcgttgcaagtgttggaagattgttcagcggtcg 371
QY 380 gactttatgcaatttaactgattggccttactgtctgctgaatggaactttgggggat 439
Db 372 gaatgtatatgttcttcgaatcgcccaattgttgcgtgctgaatgaattacggggga 431
QY 440 tccctgtttggctaaaatatgtgcctgtgatggaatttagaacaacaatcagcctttta 499
Db 432 tacctgtttgttcattatgtgccaggccacagttttccgaactgacaatgagccttca 491
QY 500 agtggtcatgcaaggatttgttcagaaaatagtaacatgataagtcagcaaaaatttgt 559
Db 492 agtatcatgacagaagtttttgacattcatagtgaacctaatgaagcaagaagacttt 551
QY 560 ttgaatctcaaggaggacacaataattatggcccagatatacaaaatgagtatgaccagtag 619
Db 552 ttgatacaaggagggtcccaataatctgtcccaggttgaaaataatgaatgagataccg 611
QY 620 aatgggaaattggtcctcctggtgaagcttatacaaaaatgggcagctcaaatggctgtag 679
Db 612 aacagttttatggggcaggggggaagttatatgacctgtgggcccgaataatggcattt 671
QY 680 gtttgaaaactggtgtcccatggtatgtgtaagcaagagagatcctctgactgtga 739
Db 672 ctcaaaatatagggtgtaccttggataatgtgccagcaatgatgctcctgactgtga 731
QY 740 ttgatactgttaatggtcttactcgcgaagggttcctccttaataagccttacaaccta 799
Db 732 tcaatactgttaattccttactgtgacaagttcataccaacactccgaacaagccca 791
QY 800 aaatggaagaaagatgagactggctgggtatacaaaaattgggtgtccaatctctcaaa 859

```
Db 792 aatgtgactgagaactgacctggtggttttaaaacttttgggctagagactcctata 851
QY 860 gaccagccgaagacattgcatcttccaggttccaggtttgttcagaaacaaatggttcattct 919
Db 852 gaccagcgggaagatattcgcttgcgttctctgcttcttcccaaaagggtggcagctac 911
QY 920 tcaattactacatgtatcatggaggaaacaaattttggccggacatcatcagggcttttca 979
Db 912 aaaactactacatgtatcatggaggaaacaaattttggctgtacatcagcgaccattta 971
QY 980 ttcaactactacatgtatgtctctctctctctgatgaataatggttctgctgaatgaaccaa 1039
Db 972 ttaccacaagtattgatgtgcgaatcgcaatcgatgtagtgcgttttaccctaggcttccaa 1031
QY 1040 agtatgggaccttgagagacttacaataagctcatcaagctatctgaaccgcttttagttt 1059
Db 1032 aatggggacaccttaagaactccataaagctatataagctgtgtgagatgcatgtctga 1091
QY 1100 catcatatgctgctggtgactagttcttgggaagttaatacaagggtctcatgtttatagatcaa 1159
Db 1092 accatgaatctatgaacatgtcacttggcccatgttgtagagccgatgtttatcacggata 1151
QY 1160 aatctggagctgtgtgctgtttttatccaaactatgactctagatattcsgtataaagctca 1219
Db 1152 attcaggagaatgtgctgcatcttctgcgaatctctgcgaacatagatgataaagaagaagttgtac 1211
QY 1220 cctttcagaataggccataaactctccatgcttccatggttccatcagcattcttcccgactgca 1279
Db 1212 agttcgaaatgtgtcttaacacttctcctgcttggctcagttagcattcttcccgactgca 1271
QY 1280 aaactccgctttacaacactgcaggttaactctcaagctcgaagctcagacataaagaatgacgc 1339
Db 1272 agaattagctataaacacgcgaaggtcggttctcagatgtctacagttcacaatggtcc 1331
QY 1340 ctgacggtggtgattctcttggcagtcatacaatgaagaacgcctactctgctga----- 1394
Db 1332 ctgagaaataaatacatcagtcagtcctcctccagaaagccgaaacctcttaaatggg 1391
QY 1395 -----tgacagggatacaccttacgctaacgagactatg----- 1427
Db 1392 aagtggttaggagaagcaggatttggggagaagccgactttgttaaaagtagttttg 1451
QY 1428 -ggaacagaaaaacgtcacagaagattcatcagactactctgtgtacatgacaaatgtaa 1486
Db 1452 tggatcatactacatactacaagaactactgactactactgtatggtatcacacaagtataa 1511
QY 1487 atatagcatctaataaggaatttctaaagaacgcaaggaaggtactctatctcactgttatgt 1546
Db 1512 ctgttggtagagatgaagattcttgaagaagggaagtaacccagttctctctattgagt 1571
QY 1547 cgcgtggtcatgtctgtcattgttctcgttcaatggaaactatcaggaactgtttatggtta 1606
Db 1572 caaagggtcagtctctcactgcttctgttgtaacaggaacttcaaggcagtgcatctggga 1631
QY 1607 cattgataatccaaacttacatcagctggcaactggaactgaaggttaagagctggttataaca 1666
Db 1632 atggatacagcttctgcctttacataataagaagcctatttctcagagcgggggaagaagc 1691
QY 1667 agatttctctcagtgcttctcgttctgttcccgcaacgcttggcgtgcatattatgatacat 1726
Db 1692 atatcaacctattaagcatgtgtcgtggtctacagaaatgctgaccatctctacga---at 1748
QY 1727 ggaatcaggagcttctaggtccagctacgtttgagcgttctcaatgaaggtcgaagaact 1786
Db 1749 ggcaggagcaggactacaagcgtcagttattgaagggtccaaacttgggacactagact 1808
QY 1787 tggcgaacagagaatggtcttacaaggttctgctgaaaggcgaatcggttaagcttcaact 1846
Db 1809 tgcctttaaataattggtcttacaagattggttgcgaggtgaacacctgggtatacaca 1868
QY 1847 ccttaagtggaagttctctctgttgtaatgggttgcaggttgcactaatgtgtcctcaaaagcagc 1906
```

```
Db 1869 accaggtagttttagtggttaaatgttgacgtcgactcgcgaagcaccgcaaacagcagc 1928
QY 1907 cctgacttggtacaaggtcacatttaacgcgcctcgagaaatgatccactagctttag 1966
Db 1929 cattgacatggtacaaggttgttgtgaccacaccccccagggtatgaacccaatcggtctg 1988
QY 1967 acatggcaagtattgggaaaaaggtcagatatggataaaatggtgaaggcgtaggtcgccatt 2026
Db 1989 acatgcttcataatgggaaaaggtctgctggttgaatggagaagaatcggaagatact 2048
QY 2027 ggcctgatacatagacaagggcagactgcagca-----aatgcagttatgctggaacgt 2080
Db 2049 ggcctagaaaaagttctgtcatgatacttgttccaaagaatgcgactacagaggaact 2108
QY 2081 tcaacgagaagaagtgcggaactaaactgcggaacacacttctcagagatggttaccatgttc 2140
Db 2109 tcttcccggaacaaatgtcttacgggatgtggagatccaaacgaagatggtaccatgtcc 2168
QY 2141 caagctgctgctgaaacacacacagtggaacactgttagtagtatttcgaagaatgggaggtta 2200
Db 2169 ctggttctgttccaaacccgtctggaacagtttgggttatattcagaggaaatgggtgaa 2228
QY 2201 atcc 2204
Db 2229 accc 2232

RESULT 12
AAZ29344
ID AAZ29344 standard; cDNA; 2972 BP.
XX
AC AAZ29344;
DT 29-FEB-2000 (first entry)
XX
DE Tomato beta galactosidase-7 cDNA.
KW Tomato beta galactosidase-7; TBG-7; clone pZBG2-1-4; fruit ripening;
Rutgers tomato plant; pectin; biofilm; gene mapping; ds.
XX
OS Lycopersicon esculentum.
XX
FH Key
FD 104..2716
FT CDS
FT /*tag= a
FT /product= "Tomato beta galactosidase-7"
FT sig_peptide 104..208
FT /*tag= b
FT mat_peptide 209..2713
FT /*tag= c
XX
PN WO964564-A1.
XX
PD 16-DEC-1999.
XX
PF 08-JUN-1999; 99WO-US12697.
XX
PR 09-JUN-1998; 98US-0088805.
XX
PA (USDA ) US DEPT OF AGRICULTURE.
XX
PI Gross KC, Smith DL;
DR WPI; 2000-097532/08.
DR P-PSDB; AAY44309.
XX
PT New beta-galactosidases, used to prepare transgenic plants with altered
fruit ripening
XX
PS Claim 2; Fig 2; 85pp; English.
XX
CC The present sequence encodes tomato beta galactosidase-7 (TBG-7). This
is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from
```


Db 993 caacctaaatggggccatttgaagatctgcatgctgtatataaaagcttttgtaaccagct 1052
Qy 1093 ttagtttcatcatat---gctgcygtgactagtcttgaagtaatacaagaggctcatgtt 1149
Db 1053 cttgttgcgtgctgactaacctcagtataataaactgggaccaaagaggagcacatgtc 1112
Qy 1150 tatagatacaaaatct-----gagccttgctgt 1176
Db 1113 tctcgtggaactcaacaacaaattggccaatatatgtccttaaatgaaggcatatgcga 1172
Qy 1177 gcttttttatccaactatgactctagatattcagtaaaagtcacacttccagaaataggcca 1236
Db 1173 gcattttatgcaaatattgatgaacatgaatcagcaacagtgaaattttaaggccaag 1232
Qy 1237 tacaatcgtccatcctggttccatcagcatttcccgactgcacaaactgcogtttacaac 1296
Db 1233 ttcaactttacctccatggf-cagtgtgtattctgccagattgcagaaatacacagcttcaac 1291
Qy 1297 actgcacaggttaactctcaagctcagacataaaagatgagcctgcaggtggtgattc 1356
Db 1292 acagctaaagggtggggccaaacttcaatacaaaaacagtgggctcagattctgttcagtt 1351
Qy 1357 tcttgccagtcatacaat----- 1374
Db 1352 ggggaataattcttcttctacaagtataactaaaagcaagctcggaagtcttttca 1411
Qy 1375 -----gaagaaacccctactgctgatgacagcgatcacacttacaagctaa 1418
Db 1412 atcttggtgacattgaaggccacttggtgtgggtgacaagaatttcaacttca 1471
Qy 1419 cggactatggaacagaaaacgtcacaaagagattcacagactatctgtgtacatgac 1478
Db 1472 aggaatactgagcaactgtaagtgcacaaagacagctctgtacactgtgtgtctgac 1531
Qy 1479 aaatgaaatagcatctaatgaaggatttctaaagcaacggaaggaacttcttctac 1538
Db 1532 caggataataattctgtatgatcatctcattttgggaggaataatgatgtagctcaac 1591
Qy 1539 tgttatgctcgcgtggtcatgtcttgcatgttttgcgtcaatggaacactatcagaaactgt 1598
Db 1592 aatt-----gatattgtagcatgctgtgattttgtgttcgcatttttgaatgggcagct 1645
Qy 1599 ttatggtacattgataatcaaaactacatacagtggtgcaactggaagttaagactgg 1658
Db 1646 tgcaggttagtgtgaaggcaaatggaatggaatggttcaacactgttaagctggttcagg 1705
Qy 1659 tattaacaagaattctctgcagtttccgttggctcccgaaacgttggcgtgcatta 1718
Db 1706 atacaacacatactgctattatctgagacggtgggattgcagaattatgtgcttctt 1765
Qy 1719 tgatactggaatcaggagttctaggtccagtcacgttggcggtctcaatgaagggtc 1778
Db 1766 ggagaaggatggggcaggttttaaggctcagataaagcttaacaggtgcaaaagcgggga 1825
Qy 1779 aagaaactggcgaacagaaatggtctcaaggttggtctgaaagcgcaactgctaaag 1838
Db 1826 tataactcaacacatctttatggacctaccaggtggggcttgaagcgaattctctgga 1895
Qy 1839 tcttactccttaagtgagggttcttcttgaatgggttcgaggttcaactaatgggtcga 1898
Db 1886 agtatatgatgtaactgaagtgaagtcaggtgaggtgactgagtttccactggtacaac 1945
Qy 1899 aagcagccccctgacttggtacagggctacattttaacgcgcctggagaaatgatcaact 1958
Db 1946 tccgtcagctcttctcgttggtacagaaagtttgatgccccagcgcggaacagatccagt 2005
Qy 1959 agctttagacatgcgaagtatggaaaaggtcagatatggaataaattggtgaagcgttagg 2018
Db 2006 tgccttgattttagtagcagtggaagaggtcaggcatgggttaattggccaccatgtagg 2065
Qy 2019 tcgcccattggcctggatcacatagca---caaggcgactgcagcaaatgcagttatgctgg 2075
Db 2066 aagatatggactttggttgaccacaaataatggtatgggaagaactgtgattatcgtgg 2125

Qy 2076 aacgttcaacagagaagaagtgcggaactaactgcggaacactctcagagatggtacca 2135
Db 2126 tgcctaccactctgataaattgtagcaaaactgtgagagattactcagcgtgtacca 2185
Qy 2136 tgcctccacgactgctgcggaacaaagtggaaacttgttagtagttagtgcgaagaatgggg 2195
Db 2186 tatactagatcgtgcggaagacatttaataatgtactagtattcttgaagaacaga 2245
Qy 2196 aggtaaatcaacaggaatttctctagtcaggagatcaagataaagaactcg 2246
Db 2246 taaaactcgttgatatttccatttctacgcgttctactgaaccatttg 2296

RESULT 14

AAZ29343
ID AAZ29343 standard; cDNA; 749 BP.

XX AAZ29343;

AC AAZ29343;

XX 29-FEB-2000 (first entry)

XX Tomato beta galactosidase-6 cDNA.

XX Tomato beta galactosidase-6; TBG-6; clone pZBG2-1-4; fruit ripening;

XX Rutgers tomato plant; pectin; biofilm; gene mapping; ds.

XX Lycopersicon esculentum.

XX Key Location/Qualifiers

XX mat_peptide 1.747

XX /tag= a

XX /product= "Tomato beta galactosidase-6"

XX /note= "no stop codon given"

XX WO9964564-A1.

XX 16-DEC-1999.

XX 08-JUN-1999; 99WO-US12697.

XX 09-JUN-1998; 98US-0088805.

XX (USDA) US DEPT OF AGRICULTURE.

XX Gross KC, Smith DL;

XX WPI: 2000-097532/08.

XX P-PSDB; AAY44308.

XX New beta-galactosidases, used to prepare transgenic plants with altered

XX fruit ripening

XX Claim 2; Fig 2; 85pp; English.

XX The present sequence encodes tomato beta galactosidase-6 (TBG-6). This

XX is a cDNA insert of clone pZBG2-1-4 from a cDNA library derived from

XX breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants.

XX TBG cDNA has been expressed in E.coli strain XL1blue MR. This is used

XX for modifying cell wall metabolism and controlling ripening of fruit by

XX altering activity of beta galactosidase II protein. Pectin with reduced

XX galactosyl content is produced for use in biofilms or solutions.

XX Fragments of the cDNA are used for gene mapping and for detecting

XX expression of beta galactosidase in plant tissue. Transgenic plants

XX with altered fruit ripening are produced by introducing DNA constructs

XX comprising a TBG cDNA associated with promoter/enhancer elements.

XX Sequence 749 BP; 205 A; 147 C; 186 G; 211 T; 0 other;

Qy Query Match

Best Local Similarity 15.2%; Score 388.4; DB 21; Length 749;

Matches 537; Conservative 0; Mismatches 211; Indels 2; Gaps 2;

Qy	274	attgaaacttatgttttctggaatggacatgagccttctctctggaataataatttgaa	333	XX	25-FEB-2000;	2000EP-0301439.
Db	1	atccagacatatgtttttggaatgttcacatgagccttctctctggaataatttgaa	60	PF		
Qy	334	ggaagatatgatctgttagattcaataaattgacaaagagacgacctttatgtcaat	393	PR	25-FEB-1999;	99US-0121825.
Db	61	ggaagatatgacctggtgaggtttgtaaaacgattcagaagcggcgtgtatgctcat	120	PR	05-MAR-1999;	99US-0123180.
Qy	394	ttacgtattggcccttaccgtctgctgaatggaactttggggatccctgtttgcta	453	PR	09-MAR-1999;	99US-0125788.
Db	121	cttcgaattggcccttaccgtttgtcagagtggaattttggagggtttccagttatgctg	180	PR	25-MAR-1999;	99US-0126264.
Qy	454	aaatatgtcgtggtatggaattagacaacaaatcagcctttttaagggtggcttatgcaa	513	PR	29-MAR-1999;	99US-0126785.
Db	181	aagtatgaactggcattagcttcagagctgataatgaacctttcaagaacgcaatgaaa	240	PR	01-APR-1999;	99US-0127462.
Qy	514	ggatttttcagaaaaatagtcacacatgatgaagtcagaaaaatttgttgaatctcaaggga	573	PR	06-APR-1999;	99US-0128234.
Db	241	gggtatgctgagaaaaattgttaacttgatgaag-atcataatcttttcgagctcaggg	299	PR	08-APR-1999;	99US-0128714.
Qy	574	ggaccataattatggccacagatcaaaaatgagtagtgaccagtagaatgggaattgg	633	PR	16-APR-1999;	99US-0129845.
Db	300	ggtcccaatactactcaccagattgagaatgagtagtggtggcctcaagccaaggttacttgg	359	PR	19-APR-1999;	99US-0130077.
Qy	634	gctcctgttaaagcttatatacaaaaatggcagctcaaatggcgttagg-tttgaaaaactgg	692	PR	21-APR-1999;	99US-0130449.
Db	360	gcacgggacatcagtatccaatggctgcaaatatggcagttgatttgaacacagg	419	PR	23-APR-1999;	99US-0130510.
Qy	693	tgctccatgatcatgtgtaagaagaaggatgctcctgatcctctgattgatacttggtaa	752	PR	28-APR-1999;	99US-0130891.
Db	420	cgtcccaatggtgatgtgcaagggaagaatgcaaccagatcctgtgatcaacacatgcaa	479	PR	30-APR-1999;	99US-0131449.
Qy	753	tgcttctactggaagggttcgtcctaataagccttacaacaaacttaaatgtgacaga	812	PR	04-MAY-1999;	99US-0132484.
Db	480	tggtttctactgtagaatttctcccaaaacacacacaaactgcaattttggactga	539	PR	06-MAY-1999;	99US-0132486.
Qy	813	agtatgactggctgtgtatcgaattcgtgtgttcccaattcctcaagacccagccgaaga	872	PR	07-MAY-1999;	99US-0132487.
Db	540	agcttggagtggaatggttctcgaaatttggcgtcccccttcacagagaccagttcagga	599	PR	11-MAY-1999;	99US-0132863.
Qy	873	cattgcaatttcagttgccaggtttgttcagaaacaattggttcattctcctaattactacat	932	PR	14-MAY-1999;	99US-0134218.
Db	600	tttggaatttctgtgttcccaatttatcaaaaggaggaggtcttttgttaactattacat	659	PR	14-MAY-1999;	99US-0134219.
Qy	933	gtacatgagggaacaaattttgccgacatcatcagggccttttctcattgcaactagcta	992	PR	14-MAY-1999;	99US-0134221.
Db	660	gtaccatgggggacgaactttggacgcactgcgggtggggccattcactaccagcta	719	PR	18-MAY-1999;	99US-0134370.
Qy	993	cgattatgctcctctcgatgaatatgg 1022		PR	19-MAY-1999;	99US-0134768.
Db	720	tgattatgcccccttcgacgagtgg 749		PR	20-MAY-1999;	99US-0134941.
RESULT	15			PR	21-MAY-1999;	99US-0135124.
AAC40074				PR	24-MAY-1999;	99US-0135623.
ID	AAC40074	standard; DNA; 1969 BP.		PR	25-MAY-1999;	99US-0136021.
XX				PR	27-MAY-1999;	99US-0136392.
AC	AAC40074;			PR	28-MAY-1999;	99US-0136782.
XX				PR	01-JUN-1999;	99US-0137222.
XX	17-OCT-2000	(first entry)		PR	03-JUN-1999;	99US-0137528.
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 26952.			PR	04-JUN-1999;	99US-0137502.
XX				PR	07-JUN-1999;	99US-0137724.
XX				PR	08-JUN-1999;	99US-0138094.
KW	Hybridisation assay; genetic mapping; gene expression control;			PR	10-JUN-1999;	99US-0138540.
KW	protein identification; signal transduction pathway;			PR	14-JUN-1999;	99US-0138847.
KW	metabolic pathway; promoter; termination sequence; ss.			PR	16-JUN-1999;	99US-0139452.
OS	Arabidopsis thaliana.			PR	17-JUN-1999;	99US-0139453.
XX				PR	18-JUN-1999;	99US-0139455.
XX	EP1033405-A2.			PR	18-JUN-1999;	99US-0139456.
XX				PR	18-JUN-1999;	99US-0139457.
XX				PR	18-JUN-1999;	99US-0139458.
XX				PR	18-JUN-1999;	99US-0139459.
XX				PR	18-JUN-1999;	99US-0139460.
XX				PR	18-JUN-1999;	99US-0139461.
XX				PR	18-JUN-1999;	99US-0139462.
XX				PR	18-JUN-1999;	99US-0139463.
XX				PR	18-JUN-1999;	99US-0139750.
XX				PR	18-JUN-1999;	99US-0139763.
XX				PR	21-JUN-1999;	99US-0139817.
XX				PR	22-JUN-1999;	99US-0139899.
XX				PR	23-JUN-1999;	99US-0140353.
XX				PR	23-JUN-1999;	99US-0140354.
XX				PR	23-JUN-1999;	99US-0140695.
XX				PR	28-JUN-1999;	99US-0140823.
XX				PR	29-JUN-1999;	99US-0140991.
XX				PR	30-JUN-1999;	99US-0141287.
XX				PR	01-JUL-1999;	99US-0141842.
XX				PR	01-JUL-1999;	99US-0142154.
XX				PR	02-JUL-1999;	99US-0142055.
PD	06-SEP-2000.			PR	06-JUL-1999;	99US-0142390.

Qy 2027 ggctggatacatagcacaggcgactgcagcaaatgcagttatgctggaacgttcaacg 2086
||||| ||| ||| ||| ||||| || ||||| ||||| ||||| |||||
Db 483 ggctgcataataagcaggttggttcttgacgcaggtgttcttatactggaacattcagag 542
Qy 2087 agaagaagtgcggaactaaactgcggaacaaccttctcagagatggtaccatgttccacgat 2146
||| ||||| || ||||| ||| ||||| ||||| ||||| ||||| |||||
Db 543 aggaagaagtgttaagaacttggaagggttctcacaagatggtaccatgtcccaaggt 602
Qy 2147 cgtggtcgaacccaagtggaaacttgtagtagtatttcgaagaatgggaggtaatccaa 2206
||||||| ||||| ||||| ||| ||| ||| ||| ||| ||||| |||||
Db 603 cgtggctaaaaccaagtggcaatctatggttcttcttgaggagtgggagagaccoga 662
Qy 2207 caggaatttctctagtcaggaga 2229
||||| ||| ||||| |||||
Db 663 acggaatctcgttggtcagaaga 685

Search completed: June 25, 2002, 06:29:18
Job time: 10523 sec

THIS PAGE BLANK (USPTO)

Db	11	ILLFSCI---FSAASASVSDHKAIIINGQRIILISGSIHYPRSTPEMWPDLIOKADG	67
Qy	67	GLDVIETVFWNGHSPGKYNFEGYDLVRFIKWQVAGLVNLRIGPYVCAEWNFGGF	126
Db	68	GLDVIQTVFWNGHSPGNYFEERYDLVRFIKVQEGSLVNLRIQPYVCAEWNFGGF	127
Qy	127	PWLKIVYVGMERTNQQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPVE	186
Db	128	PWLKIVYVGMERTNQQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPVE	187
Qy	187	WEIGAPGKAYTKAAQMAVGLTGVPMCKQEDADPDVIDFCNGFYCEGRRPNPKYPK	246
Db	188	WEIGAPGKAYTKAAQMAVGLTGVPMCKQEDADPDVIDFCNGFYCEGRRPNPKYPK	247
Qy	247	MWETVMTGMYTKFGGPIQFPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLFI	306
Db	248	MWETVMTGMYTEFGGAVPTPAEDIAFVAREVQNGSFFNYMHGTTNFGRTAGGPFM	307
Qy	307	ATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNOEAHVRSK	366
Db	308	ATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNOEAHVRSK	367
Qy	367	SGACAAFLSNYSRYSVKTQFQNRPNLPPWISILPDKCTAVYNTAQVNSOSSIKWTP	426
Db	368	SD-CAAEFLANVDKYSVKGSGGQYDLPWISILPDKCTAVYNTAQVNSOSSIKWTP	426
Qy	427	AGGSLWSQSYNETPTADDSDTLTANGLEQKNVTRDSSDYLWYMTNVNIASNEGFLKNG	486
Db	427	VHSGFPWQSFIEETSSDETDTLLDGLYEQLNITRDYILWYMTDITIGSDEAFKNG	486
Qy	487	KDPYLTVNSAGHLVHFVNGKLSGTVYGLDNPDKITYSGNKLKRAKINKILLSVSVGLP	546
Db	487	KSPLLTIFSAGHALVNFINGQLSGVYGSLENPKLFSQNVNLRSGINKLALLSVGLP	546
Qy	547	NGVGHYDTWAGVGLPVTLSGLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVHR	606
Db	547	NGVGHYDTWAGVGLPVTLSGLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVHR	606
Qy	607	GSLMAQKPLTWYKATFNAGNDPLALDMSMGKQIWIINGEGYGRHWPGYIAGDCSK	666
Db	607	GSLMAQKPLTWYKATFNAGNDPLALDMSMGKQIWIINGEGYGRHWPGYIAGDCSK	666
Qy	667	CSYAGTFNEKKQTCNGQPSQWYHVPWSLWPKPSNLLVWFEWGGNTGTSILVSR	722
Db	667	CSYAGTYDDKCRTHCEGSPSQWYHVPWSLWPKPSNLLVWFEWGGNTGTSILVSR	722
RESULT 2			
BDAL_LYCES	STANDARD:	PRT:	835 AA.
AC	BGAL_LYCES		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	01-FEB-1996 (Rel. 33, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Beta-galactosidase precursor (EC 3.2.1.23) (Lactase) (EXO-(1->4)-beta-D-galactanase).		
OS	Lycopersicon esculentum (Tomato).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; euasterids I; Solanales; Solanaceae; Solanum.		
OX	NCBI_taxid=4081;		
RP	[1]		
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.		
RC	STRAIN=CV. AILSA CRAIG; TISSUE=pericarp;		
RX	MEDLINE=95357407; PubMed=7630937;		
RA	Carey A.T., Holt K., Picard S., Wilde R., Tucker G.A., Bird C.R.,		
RA	Schuch W., Seymour G.B.;		
RT	"Tomato exo-(1->4)-beta-D-galactanase. Isolation, changes during		
RT	ripening in normal and mutant tomato fruit, and characterization of a		
RT	related cDNA clone."		
RL	Plant Physiol. 108:1099-1107(1995).		
CC	-1- FUNCTION: INVOLVED IN CELL WALL DEGRADATION. DEGRADATES		
CC	POLYSACCHARIDES CONTAINING BETA-(1->4)-LINKED GALACTANS, ACTING		

CC	AS AN EXO (1->4)-BETA-D-GALACTANASE. HAS A PH OPTIMUM OF 4.5.		
CC	-1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-		
CC	galactose residues in beta-D-galactans.		
CC	-1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.		
CC	-1- SIMILARITY: CONTAINS 1 SUEL-TYPE LECTIN DOMAIN.		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL: X83854; CAA58734.1; -		
DR	InterPro: IPR000922; Gal_Lectin.		
DR	InterPro: IPR001944; Glyco_hydro_35.		
DR	Pfam: PF02140; Gal_Lectin; 1.		
DR	Pfam: PF01301; Glyco_hydro_35; 1.		
DR	PRINTS: PR00742; GLHYDRLASE35.		
DR	ProDom: PD005612; Gal_Lectin; 1.		
DR	PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.		
DR	PROSITE: PS02028; SUEL_Lectin; 1.		
KW	Hydrolase; Glycosidase; Signal.		
FT	SIGNAL 1 22		
FT	CHAIN 23 835		
FT	DOMAIN 749 835		
FT	ACT_SITE 180 180		
FT	ACT_SITE 249 249		
FT	NUCLEOPHILE (POTENTIAL).		
FT	SEQUENCE 835 AA; 93336 MW; 94C9685F95C4A646 CRC64;		
Qy	Query Match	74.5%; Score 2922.5; DB 1; Length 835;	
Qy	Best Local Similarity	72.2%; Pred. No 5,1e-196;	
Qy	Matches 518; Conservative	94; Mismatches 102; Indels 3; Gaps 2;	
Qy	6	VLLLVICLLDFFSSKASVSYDDRAIIINGKRLKILISGSIHYPRSTPEMWPDLIOKAD	65
Db	7	MLLMLLLCL--VWSGSIASVSDHKAIIINGQRIILISGSIHYPRSTPEMWPDLIOKAKE	64
Qy	66	GLDVIETVFWNGHSPGKYNFEGYDLVRFIKWQVAGLVNLRIGPYVCAEWNFGG	125
Db	65	GGVDVIQTVFWNGHSPGKYNFEGYDLVRFIKWQVAGLVNLRIGPYVCAEWNFGG	124
Qy	126	FPVWLKIVYVGMERTNQQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPV	185
Db	125	FPVWLKIVYVGMERTNQQPKVAMQGFVQKIVNMKSENLFESQGGPITMAQIENYGPV	184
Qy	186	EWEIGAPGKAYTKAAQMAVGLTGVPMCKQEDADPDVIDTCNGFYCEGRRPNKYPK	245
Db	185	EWEIGAPGKAYTKAAQMAVGLTGVPMCKQEDADPDVIDTCNGFYCEGRRPNKYPK	244
Qy	246	KWTEVMTGMYTKFGGPIQFPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLF	305
Db	245	KWTEVMTGMYTKFGGPIQFPAEDIAFVAREVQNGSFFNYMHGTTNFGRTSSGLF	304
Qy	306	IATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNOEAHVRS	365
Db	305	IATSYDYDAPLDYGLLNEPKYGHLDLHKAIKLSEPALVSSVAATVSLGSNOEAHVRS	364
Qy	366	KSGACAAFLSNYSRYSVKTQFQNRPNLPPWISILPDKCTAVYNTAQVNSOSSIKMT	425
Db	365	ESGACAAFLSNYSRYSVKTQFQNRPNLPPWISILPDKCTAVYNTAQVNSOSSIKMT	424
Qy	426	PAGGSLWSQSYNETPTADDSDTLTANGLEQKNVTRDSSDYLWYMTNVNIASNEGFLKN	485
Db	425	PVSRGFSWESNEDEAASHED-DTFTVWGLLEQINITRDYILWYMTDIEDTPEGFLNS	483
Qy	486	GKDPYLTVNSAGHLVHFVNGKLSGTVYGLDNPDKITYSGNKLKRAKINKILLSVSVGL	545
Db	484	GNWPLTVFSAGHALVNFINGQLSGVYGSLENPKLFSQNVNLRSGINKLALLSVGL	543
Qy	546	PNVGHYDTWAGVGLPVTLSGLNEGRNLAKQKWSYKVLKGESLSLSLSSSSSVVW	605

```

||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
544 PNVGPHETWAGLVGVSLSNGLNEGRDLTWQKWFYKVGKGLGALSLSHSGSPSVEWY 603
QY RGSLSMAOKQPLTWYKATFNAGPGNDPLALDMSKGOIWINGEGVGRHWPYIAQGDPCS 665
Db EGSLSVAOKQPLSWTKTTFNADGNEPLALDMSKGOIWINGEGVGRHWPYIAQGDPCS 663
QY KCSYAGTFNEKKCOTKNCQSPQRWYHVPWSLKPESGNLLVVFEEWGGNPTGISLVR 722
Db VCNVTGWFEDEKKCLTNGEGSGQRWYHVPWSLWLYPTGNLLVVFEEWGGDPYGIILVK 720

RESULT 3
ID BGAL_ASPOF STANDARD; PRT; 832 AA.
AC P45582;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase).
OS Asparagus officinalis (Garden asparagus).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Asparagaceae;
OC Asparagus.
OX NCBI_TaxID=4686;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LIMBRAS 10; TISSUE=Sphear;
RX MEDLINE=95303968; PubMed=7784512;
RA King G.A., Davies K.M.;
RT "Cloning of a harvest-induced beta-galactosidase from tips of
RL Plant Physiol. 108:419-420(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: BELONGS TO SUPERFAMILY 1 SUEL-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: X77319; CAA54525.1;
DR InterPro; IPR000922; Gal_lectin.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF02140; Gal_Lectin; 1.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRLASE35.
DR ProDom; PD005612; Gal_lectin; 1.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
DR PROSITE; PS02228; SUEL_LECTIN; 1.
KW Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 832 BETA-GALACTOSIDASE.
FT DOMAIN 741 832 SUEL-TYPE LECTIN.
FT ACT_SITE 183 183 PROTON DONOR (POTENTIAL).
FT ACT_SITE 252 252 NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 832 AA; 92213 MW; 94ABDC61EC4164AE CRC64;

Query Match 73.08; Score 2863; DB 1; Length 832;
Best Local Similarity 71.48; Pred. No. 7.2e-192;
Matches 514; Conservative 96; Mismatches 102; Indels 8; Gaps 5;

QY 6 VLLLVICLIDFFS--SVKASVYDDRAIINGKRKILSGSIHYPRSTPQWPDLIQKA 63
Db 6 VLMLVALLAAVPPAVTVDHKSVIINGQRRIISGSIHYPRSTPQWPDLIQKA 65
QY 64 KDGGLDVIETVYVFWNGHEPSPGKYNFEGRYDLVRFLKMWQVAGLYVNLRIGPYVCAEWNF 123

```

```

||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
66 KDGGLDVIETVYVFWNGHEPSPGKYNFEGRYDLVRFLKMWQVAGLYVNLRIGPYVCAEWNF 125
QY GGFVWLKLYVPGMEFRTNNQPKFVAMOGFVOKIVNMKSENLFSOGGPIIMAGIENEYG 183
Db GGFVWLKLYVPGMEFRTNNQPKFVAMOGFVOKIVNMKSENLFSOGGPIIMAGIENEYG 185
QY PVEWEIGAPGKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPY 243
Db PVEYDGAAGKSYTNWAAKMAVGLKTVGPWIMCKQEDAPDVIDTCNGFYCEGFRPNKPY 245
QY KPMWTEVMTGWYTKFGGPIIPQPAEDIAFSVARFVQNGSFFNYMYHGTGNTGRTSSG 303
Db KPMWTEVMTGWYTKFGGPIIPQPAEDIAFSVARFVQNGSFFNYMYHGTGNTGRTSSG 305
QY LFIATSYDYDAPLDEYGLLNEPKVGHLDLHKAIKLSPALVSVAAYTSLSGNOEAVHY 363
Db PFISTSYDYDAPLDEYGLLNEPKVGHLDLHKAIKLSPALVSVAAYTSLSGNOEAVHY 365
QY RSKGACAAFLSNYDSRYSVKVTFTQNPYNLPWWSISILPDKCTAVYNTAQVNSQSSIK 423
Db RSKS--SCAAFLANFSRYIATVTENGHYNLPWWSISILPDKCTAVYNTARVGAQTTMK 424
QY MTPAGGWSQSYNEETPTADSDTLFANGLWEOKNVTROSDSYLWYNTVNIASNEGFL 483
Db MQYL--GGFSWKAYTETDNLND--NTFTKDLGVEQLSTWDRSDYLYWYTYVDIAKNEEFL 482
QY KNGKDPYLYTVMSAGHVLHGVNGKLSGTVYGLDNLNPKLTVSGNVKLRAGINKISLSVSV 543
Db KTGKPYLYTVMSAGHVLHGVNGKLSGTVYGLDNLNPKLTVSGNVKLRAGINKISLSVSV 542
QY GLPNVGHYDWTNAGVILGPVTLGSLNEGRNLAKQKSYKVGLKAGESLSLSHSGSSVE 603
Db GLPNVGNHFEFTWNTGLVGLPVTGLNEGRDLSLQKWTYQIGLHGETLSLSHSGSSNVE 602
QY WVRGSLMAQKOPLTWYKATFNAGPGNDPLALDMSKGOIWINGEGVGRHWPYIAQGD 663
Db WGEAS---OKOPLTWYKATFNAGPGNDPLALDMSKGOIWINGEGVGRHWPYIAQGD 659
QY CSKSYAGTFNEKKCOTKNCQSPQRWYHVPWSLKPESGNLLVVFEEWGGNPTGISLVR 723
Db CGSCDYRGTYNEKKCLSNCGEASQRWYHVPWSLKPESGNLLVVFEEWGGNPTGISLVR 719

RESULT 4
ID BGAL_DIACA STANDARD; PRT; 731 AA.
AC Q00662;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative beta-galactosidase precursor (EC 3.2.1.23) (Lactase)
DE (SRL2 protein).
DE CARS12.
OS Dianthus caryophyllus (Carnation) (Clove pink).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllales; Caryophyllaceae; Dianthus.
OX NCBI_TaxID=3570;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. WHITE SIM; TISSUE=Petal;
RX MEDLINE=91329738; PubMed=1868223;
RA Raghothama K.G., Lawton K.A., Goldsbrough P.B., Woodson W.R.;
RT "Characterization of an ethylene-regulated flower senescence-related
RL gene from carnation."
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- TISSUE SPECIFICITY: SENESCING FLOWER PETALS.
CC -1- INDUCTION: BY ETHYLENE.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----

```


Qy 181 EYGPWEIEGAPKAYTKWAAQMAVGLKTPWIMCKQEDAPDPVIDTCNGFYCEGFRPN 240
 Db 182 EYGNVSSYGAEGRAYIDWCANMNSLDIGVPMCKQEDAPDPVIDTCNGFYCEGFRPN 242
 Qy 241 KPYKPKMTWNTGTYTFFGPIQORPAEDAFSAFVAREVQNGSFNFYMYHGGTNGFR 300
 Db 243 NPSPGKMTWNTGTYTFFGPIQORPAEDAFSAFVAREVQNGSFNFYMYHGGTNGFR 302
 Qy 301 SSGFIATSYDAPLDEYGLNPKYGLHDLKAKLSEPALVSSAAVTSLSGNOEA 360
 Db 303 AGGPVITTSYDAPLDEYGLNPKYGLHDLKAKLSEPALVSSAAVTSLSGNOEA 362
 Qy 361 HVRKSGAACAAFTSNYSRVSKVTFQNRPNLPPWSISILPCKTAVYNTAQVNSOSS 420
 Db 363 TVY-STNEKSSCFIGNVATADALVNFKGDYVNPVPAWSVSLPDCDKAYNTARVNTQTS 421
 Qy 421 SIKM-TPAGGGLSWQ-SYNEETPTADSDTLTANGLWEQKNVTRDSSDYLTWMTN 475
 Db 422 IITEDSCDEPKLWTRPEFTTQRTILKSGDLIAKGLVDKQVNDASDYLTWMTN 481
 Qy 476 IASNEGFLKNGKDPY-LTVMSAGHVLHVFVNGKLSGTGYGTLDNPKLTSYGNV 529
 Db 482 L-----DKKDPWISRNMSLVHSNAHLVHAYVNGKYVGNQIVRDKNKDYREPK 533
 Qy 530 RAGINKISLVSGLPNVGHYDTWNAAGVLPVTLGSLNEG-SSRLAKQKWSYKVL 586
 Db 534 VHGTHLALLSVSGLQNGYFPFESGPTGNGPVKLVGKDETEKDKLSKHQWDYKIG 593
 Qy 587 KGESLSLHSL-SGSSSVWVRGSLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQ 644
 Db 594 NGFNHLKFSMSAGHHHRKSTELKPADRM-LSWYKANFKAPLGDGPVIVDLNGL 652
 Qy 645 WINGEGVRHWPYGA-OGDCSCSVAGTNEKRCQNGOPSORVHVRSLKPSG- 701
 Db 653 WINGQISGRYPFNSNSDEGCTEEDYRGEYSDKCAFCWCKPQRTWYHVRSLN 712
 Qy 702 NLLVVFEEWGNPT 715
 Db 713 NTITLFEEMGDPS 726

RESULT 6

RGAL_XANMN STANDARD; PRT; 598 AA.
 AC P48982;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase).
 GN BGA.
 OS Xanthomonas manihotis.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xanthomonas.
 OX NCBI_TaxID=43353;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-38.
 RC STRAIN=7AS1;
 RX MEDLINE=96121416; PubMed=8563148;
 RA Taron C.H., Benner J.S., Hornstra L.J., Guthrie E.P.;
 RT "A novel beta-galactosidase gene isolated from the bacterium
 RT Xanthomonas manihotis exhibits strong homology to several eukaryotic
 RT beta-galactosidases."
 RL Glycobiology 5:603-610(1995).
 CC -1- FUNCTION: PREFERENTIALLY HYDROLYZES BETA(1->3) GALACTOSYL LINKAGES
 CC -1- OVER BETA(1->4) LINKAGES. HAS A PH OPTIMUM OF 4.5.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
 CC galactose residues in beta-D-galactosides.
 CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L35444; AAC41485.1; -
 DR InterPro: IPR001944; Glyco_hydro_35.
 DR Pfam: PF01301; Glyco_hydro_35; 1.
 DR PROSITE: PS01182; GLYCOSYL_HYDROL_F35; 1.
 KW Hydrolase; Glycosidase; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 598 BETA-GALACTOSIDASE.
 FT ACT_SITE 184 184 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 260 260 NUCLEOPHILE (POTENTIAL).
 SQ SEQUENCE 598 AA; 66085 MW; DB4C3F05E01435BF CRC64;
 Query Match 13.4%; Score 525.5; DB 1; Length 598;
 Best Local Similarity 26.0%; Pred. No. 3.5e-29;
 Matches 191; Conservative 90; Mismatches 279; Indels 175; Gaps 26;
 Qy 1 MLRTNVLVLLVLCILDFESSVKASVSY-----DDRAIIINGKRKILISGSIHYPRSTPQM 55
 Db 1 MLRTTAPLVLALALAPAAATPESWTFGTGTFQFVRDCKPYQLLSGAIHFQIRPRAY 60
 Qy 56 WPDLIQAKDGLDVIETTYVFWNGHEPSPGKYNFEGRYDLVRFIRKVRQAGLYVNLRTGP 115
 Db 61 WKDLQKARALQNTVETTYVFWNLVEPQGGQFDFSGNNDVAFAVKEAAQGLNVLTRPGP 120
 Qy 116 YVCAEWNEGGPVLWKLYVPNGHEFRTNQPFKVMQGFQKIVNMKSENLENSESQGGPIIM 175
 Db 121 YACAEWAGGYPVWLFGKGNIRVRSRDRFLAASQAYLDALAKQV--QPLLNNHNGGPIIA 178
 Qy 176 AQIENEYGPVEWIEGAPGKAYTKWAAQMAVGLKTPWIMCKQEDAPDPVIDTCNGFYCE 235
 Db 179 VQVENEYSY-----ADHAY--MADNRAMVYKAGFKALLFTSGADML---ANGTLPD 228
 Qy 236 -----GFRPN-----KPKPKMTWNTGTYTFFGPIQORPAEDAFSAFVARE 278
 Db 229 TLAVVNFAPGEAKSAFADKLIFRPDPQPRMVGWYAGWFDHNGKP---HAATDARQAEEF 285
 Qy 279 --VQNGSFFVYMYHGGTNGFRSSGLF-----IATSYDYPADLPDEYGLNPK 326
 Db 286 EMILRQGHSAFLYMFIGTSGFGFMNGANFQNPNSDHYAPQTSYDYDAILDEAG-HPTPK 344
 Qy 327 YGHLRDLHKATKLEPVALVSSVAAVTSLSGNOEAHVYRSKGCACAAFLNSDYSRVKVT 386
 Db 345 FALMRD-----AIARVTGV----- 358
 Qy 387 FQNRPNLP-PWISILPCKTAVYNTAQVNSQSSIKMTAGGGLSWQSYNEETPTADD 445
 Db 359 ---QPPALPAPITTTTL-----ATPLRESASL-----WD--NLPTPIAD 394
 Qy 446 SDTLTANGLWEQKNVTRDSSDYLTWMTNVTNITASNEGFLKNGKDPVLTWMSAGHVLHVF 505
 Db 395 T-----PQPMQFGQDYGYILYRTTITGPR-----KGP-LYLGVDVDRVARVVD 437
 Qy 506 KGLSGTYGTLDNPKLTSYGNVVKLRAGINKISLVSGLPNVGHYDTWNAAGVLPVTL 565
 Db 438 QRPVGSVERRLQOVSL-----EVEIPAGQHTLDLVENSGRINYGTRMADGRAGLVDPVLL 493
 Qy 566 SGLNFGSNLAKQKWSYKVLKGLGESLSLHLSGSSSVWVRGSLMAQKQPLTWYKATPNA 625
 Db 494 D-----SQL-----TGQAFPLPMR-----TPDSIRGWTGKAVQGFAPFRTGLRI 534
 Qy 626 FGGNDPLALDMSMGKQIWIWNGEGVGRHWPYIAAGDCSKCSYAGTNEKCKQINCQOP 685
 Db 535 GTPPTDY-LDMRAFCKGFAWNGVNLGRHW-----NTGPTQALYLRPSRARV 580
 Qy 686 SORVYHVRSLKPS 700
 Db 581 TTRWSSSTWMLHPS 595

RESULT 7
 BGAL_HUMAN STANDARD; PRT; 677 AA.
 AC P16278;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Beta-galactosidase precursor (BC 3.2.1.23) (Lactase) (Acid beta-galactosidase).
 GN GLB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Testis;
 RX MEDLINE=90062209; PubMed=2511208;
 RA Morreau H., Galjart N.J., Gillemans N., Willemssen R., van der Horst G.T.J., D'Azzo A.;
 RT "Alternative splicing of beta-galactosidase mRNA generates the classic lysosomal enzyme and a beta-galactosidase-related protein."; J. Biol. Chem. 264:20655-20663(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90262647; PubMed=2111707;
 RA Yamamoto Y., Hake C.A., Martin B.M., Kretz K.A., Ahern-Rindell A.J., Naylor S.L., Mudd M., O'Brien J.S.;
 RT "Isolation, characterization, and mapping of a human acid beta-galactosidase cDNA."; DNA Cell Biol. 9:119-127(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=89061717; PubMed=3143362;
 RA Oshima A., Tsuji A., Negao Y., Sakuraba H., Suzuki Y.;
 RT "Cloning, sequencing, and expression of cDNA for human beta-galactosidase."; Biochem. Biophys. Res. Commun. 157:238-244(1988).
 RN [4]
 RP VARIANTS MORQUIO B LEU-273; HIS-482 AND CYS-509.
 RX MEDLINE=92026088; PubMed=1928092;
 RA Oshima A., Yoshida K., Shimmoto M., Fukuhara Y., Sakuraba H., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in morquio B disease."; Am. J. Hum. Genet. 49:1091-1093(1991).
 RN [5]
 RP VARIANTS GM1 CYS-49; THR-51 AND CYS-201.
 RX MEDLINE=91353572; PubMed=1909089;
 RA Nishimoto J., Nanba E., Inui K., Okada S., Suzuki K.;
 RT "GM1-gangliosidosis (genetic beta-galactosidase deficiency): identification of four mutations in different clinical phenotypes among Japanese patients."; Am. J. Hum. Genet. 49:566-574(1991).
 RN [6]
 RP VARIANTS GM1 THR-51; ARG-123; CYS-201; CYS-316 AND GLN-457.
 RX MEDLINE=91328151; PubMed=1907800;
 RA Yoshida K., Oshima A., Shimmoto M., Fukuhara Y., Sakuraba H., Yanagisawa N., Suzuki Y.;
 RT "Human beta-galactosidase gene mutations in GM1-gangliosidosis: a common mutation among Japanese adult/chronic cases."; Am. J. Hum. Genet. 49:435-442(1991).
 RN [7]
 RP VARIANT GM1 HIS-482.
 RX MEDLINE=93138608; PubMed=1487238;
 RA Mosna G., Fattore S., Tubiello G., Brocca S., Trubia M., Gianazza E., Gatti R., Danesino C., Mineili A., Piantanida M.;
 RT "A homozygous missense arginine to histidine substitution at position 482 of the beta-galactosidase in an Italian infantile GM1-gangliosidosis patient."; Hum. Genet. 90:247-250(1992).
 RN [8]
 RP VARIANTS GM1 CYS-208; ARG-578; HIS-590 AND GLY-632.

RX MEDLINE=94027054; PubMed=8213816;
 RA Boustany R.-M., Qian W.-H., Suzuki K.;
 RT "Mutations in acid beta-galactosidase cause GM1-gangliosidosis in American patients."; Am. J. Hum. Genet. 53:881-888(1993).
 RN [9]
 RP VARIANT GM1 MET-82.
 RX MEDLINE=94256487; PubMed=8198123;
 RA Chakraborty S., Rafi M.A., Wenger D.A.;
 RT "Mutations in the lysosomal beta-galactosidase gene that cause the adult form of GM1 gangliosidosis."; Am. J. Hum. Genet. 54:1004-1013(1994).
 RN [10]
 RP VARIANTS MORQUIO B HIS-83 AND CYS-482.
 RX MEDLINE=96049832; PubMed=7586649;
 RA Ishii N., Oohira T., Oshima A., Sakuraba H., Endo F., Matsuda I., Sukegawa K., Orii T., Suzuki Y.;
 RT "Clinical and molecular analysis of a Japanese boy with Morquio B disease."; Clin. Genet. 48:103-108(1995).
 RN [11]
 RP VARIANTS GM1 HIS-59; ASN-591 AND CYS-591.
 RA Morone A., Bardelli T., Donati M.A., Giorgi M., Di Rocco R., Gatti R., Taddeucci G., Ricci R., D'Azzo A., Zammarchi E.;
 RT "Identification of new mutations in six Italian patients affected by a variant form of infantile GM1-gangliosidosis with severe cardiomyopathy."; Am. J. Hum. Genet. 61:A258-A258(1997).
 RN [12]
 RP VARIANTS GM1 H-59; S-121; C-208; M-240 AND N-491, AND VARIANTS P-10; C-521 AND G-532.
 RX MEDLINE=99268417; PubMed=10338095;
 RA Silva C.M.D., Severini M.H., Sopelsa A., Coelho J.C., Zaha A., d'Azzo A., Giugliani R.;
 RT "Six novel beta-galactosidase gene mutations in Brazilian patients with GM1-gangliosidosis."; Hum. Mutat. 13:401-409(1999).
 CC -!- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM GALACTOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANS.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-galactose residues in beta-D-galactosides.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.
 CC -!- DISEASE: DEFECTS IN GLB1 ARE THE CAUSE OF GM1-GANGLIOSIDOSIS, AN AUTOSOMAL RECESSIVE DISORDER WITH THREE MAJOR CLINICAL PHENOTYPES THAT ARE DISTINGUISHED ACCORDING TO THE AGE OF ONSET AND SEVERITY OF SYMPTOMS: INFANTILE, JUVENILE AND ADULT. THE INFANTILE FORM IS RAPIDLY PROGRESSIVE AND RESULTS IN SEVERE CENTRAL NERVOUS SYSTEM DEGENERATION AND VISCEROMEGLY, WITH DEATH USUALLY BETWEEN THE FIRST AND SECOND YEAR.
 CC -!- DISEASE: DEFECTS IN GLB1 ARE THE CAUSE OF MUCOPOLYSACCHARIDOSIS IV B (OR MORQUIO B SYNDROME) WHICH IS CHARACTERIZED BY SEVERE BONE DEFORMITIES WITHOUT CNS INVOLVEMENT.
 CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; M27507; AAA51819.1; -;
 EMBL; M34423; AAA51823.1; -;
 EMBL; M22590; AAA51822.1; -;
 PIR; A32611; A32611.
 PIR; A31673; A31673.
 PIR; A32688; A32688.
 PIR; B37086; B37086.
 MIM; 230500; -;

CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF006749; AAB81350.1; -;
DR EMBL; AF029974; AAB86405.1; -;
DR InterPro; IPR001944; Glyco_hydro_35;
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDRASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
FT SIGNAL 1 24
FT PROPEP 25 29 BY SIMILARITY.
FT CHAIN 30 669 BETA-GALACTOSIDASE.
FT ACT_SITE 189 189 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 269 269 NUCLEOPHILE (POTENTIAL).
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 547 547 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 483
FT CONFLICT 483 R -> P (IN REF. 2).
SQ SEQUENCE 669 AA; 75229 MW; 35B84933BB5E2F76 CRC64;

Query Match 12.5%; Score 490.5; DB 1; Length 669;
Best Local Similarity 26.3%; Pred. No. 1.1e-26;
Matches 182; Conservative 91; Mismatches 271; Indels 147; Gaps 26;

QY 8 LLLVCLDLDFSSVRAS-----VSDDRALLINIKRKILISGSIHYRSTPMQMPDLIQ 61
DB 12 LLLVPLLLGARGLRNASORTKIDYGHNRFLKDGQPFRIISGSIHYRSTPMQMPDLIQ 71
QY 62 KAKDGLDVIETVFWNGHEPSPGKYNEGRVYDLVRFKTMQVAGLYNLRIGPVYCAEW 121
DB 72 KMKMAGLNAIQTVFWNHEPSPGKYNEGRVYDLVRFKTMQVAGLYNLRIGPVYCAEW 131
QY 122 NFGGPPVWLKYPVGEFFTNPPFVAMQGVQKIVNMKNLENFESGGPIIMAQIENE 181
DB 132 DMGGLPAWLLKESILRSSDPDYLAAVDKWLGVLLPKMKP--LLYQNGSPIITQVENE 189
QY 182 YGP-----VVEIGAPGRKAYTKAAQ---MAVGLKTVGVNIMCKQEDAPD 223
DB 190 YGSYFTCDYDLRFLQRRFRDLHGDVLLFTTDDGAHEKFLQCGAQG---IYATVDFGPD 246
QY 224 PYDTCNCFYCGCFRPNKPYKPKMTETVWGTWYTFKGGPIPORPAEDIAFVAREVQNG 283
DB 247 ANITA--AFQIQ--RKSEPRGLVNSEFTYGTWLDHGWGPHSRVRETVASSL-HDVLAHG 301
QY 284 SFENYMYHGTGNFGRTSGLGFI-----ATSDYDAPLDYEGLLNEPKYGHRLDLHKA-- 336
DB 302 ANVLYMFTGGTFNAYW-NGANIPYQOPTSYDYDAPLSEAGDLTD-KYFALROVIRKFE 359
QY 337 -----IKLSEPALVSSAAVTSLSGNOEHAHVYRSKSGACAFSLNYSRYSKVYTFQNR 390
DB 360 KYVEGFIPTSTPKFYAGYKVALOKLTVEDA-----LNVLCPA--GPISLXPLTITQVKQ 412
QY 391 PYNLPWSTISILPCKTAVYNTAQNVSQSSSIKMTMPAGGGLSQSNQNEETPTADSDTLT 450
DB 413 YGFVLVYRTTLQDCS-----NPTPLSSPLNGVR-----DRAYVAVDG---V 451
QY 451 ANGLWEQKNVTRDSSDYLYMTNVTNVIASNEGFLKNGKDPYLFVMSAGHYLVHVFVNGKLSG 510
DB 452 PQGVLSR-----YVITLNTIGQAGATLD-----LLVENMGRVNY-----GRVIN 491
QY 511 TVYGTLDNPKLTYSGNVKLRAGINKISLLSVSGVLPNVGVHYDTWNAGVGLVPTLSGLNE 570

DB 492 DFKGLISN--LTLGSSVLTDWMIFPLDTEDA-----VRSHLGWGH-----GRNH 533
QY 571 GSRNLAKQKWSYKVLKGSLSLHSGSSSVWVVRGSLMAQKQLPTWYKATFNAPGGND 630
DB 534 GRQD-----NKAFAHSSNYTLP-----AFYAGNFSIPSGIP 565
QY 631 PLALD-----MASMGRGQIWIWINGEGVGRHWPG 657
DB 566 DLPQDTFFIQFSGWTKGQVWINGENLGRYWP 596

RESULT 9
BGAL_MOUSE
ID BGAL_MOUSE STANDARD; PRT; 647 AA.
AC P23780;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-galactosidase precursor (EC 3.2.1.23). (Lactase) (Acid beta-
GN galactosidase).
GN GLB1 OR GLB-1 OR BGL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91076843; PubMed=2124109;
RA Nanba E., Suzuki K.;
RT "Molecular cloning of mouse acid beta-galactosidase cDNA: sequence,
RT expression of catalytic activity and comparison with the human
RT enzyme.";
RL Biochem. Biophys. Res. Commun. 173:141-148(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DEA/2J;
RX MEDLINE=91298941; PubMed=1906271;
RA Nanba E., Suzuki K.;
RT "Organization of the mouse acid beta-galactosidase gene.";
RL Biochem. Biophys. Res. Commun. 178:158-164(1991).
CC -!- FUNCTION: CLAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
CC GANGLIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCAN.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -!- SUBCELLULAR LOCATION: Lysosomal.
CC -!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M57734; AAA37293.1; -;
DR EMBL; M75122; AAA37292.1; -;
DR EMBL; M75137; AAA37292.1; JOINED.
DR EMBL; M75107; AAA37292.1; JOINED.
DR EMBL; M75108; AAA37292.1; JOINED.
DR EMBL; M75109; AAA37292.1; JOINED.
DR EMBL; M75111; AAA37292.1; JOINED.
DR EMBL; M75112; AAA37292.1; JOINED.
DR EMBL; M75113; AAA37292.1; JOINED.
DR EMBL; M75114; AAA37292.1; JOINED.
DR EMBL; M75115; AAA37292.1; JOINED.
DR EMBL; M75116; AAA37292.1; JOINED.
DR EMBL; M75117; AAA37292.1; JOINED.
DR EMBL; M75118; AAA37292.1; JOINED.
DR EMBL; M75119; AAA37292.1; JOINED.
DR EMBL; M75120; AAA37292.1; JOINED.

DR EMBL; M75121; AAA37292.1; JOINED.
DR PIR; A37086; A37086.
DR MGI; 88151; G1b1.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLHYDLASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
FT SIGNAL 1 24
FT PROPEP 25 29
FT CHAIN 30 647 BETA-GALACTOSIDASE.
FT ACT_SITE 189 189 PROTON DONOR (POTENTIAL).
FT ACT_SITE 269 269 NUCLEOPHILE (POTENTIAL).
FT CARBOHYD 27 27 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 500 500 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 504 504 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 510 510 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 557 557 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 517 517 N -> D (IN REF. 2).
FT CONFLICT 539 539 G -> R (IN REF. 2).
SQ SEQUENCE 647 AA; 73121 MW; 0E68EA66A10803A CRC64;

Query Match 12.2%; Score 479.5; DB 1; Length 647;
Best Local Similarity 26.1%; Pred. No. 6.2e-26;
Matches 176; Conservative 85; Mismatches 257; Indels 157; Gaps 23;

QY 25 VSDDDRAIIINGKRKILISIIHYPRTPQWPDLIQAKDGGDLVETVYFNGHBPSP 84
DB 35 LDYSRDLKDGDPFRYISGSIHYFRIPREYEDRLKMKMAGLNAQMTVYVNFHEPQP 94
QY 85 GKNFEGRYDLVRFKIVQAGVYVNLRIQYVCAENFGSPVWLKVVPCMEPTNQP 144
DB 95 GQFEFGDRDVEHFQIAHELGLLVILRPGYICAEWMDGMLPAWLLEKOSIVLRSDPD 154
QY 145 FKVAMQGVQKIVNMKSENLFESQGGPIIQAQIENEYGP-----VWEI 189
DB 155 YLVAVDKWLAVLPKMP--LLYQNGGPIITVQVENEYGYFACDYDLRFVHFRYHL 212
QY 190 GARGKATKWAQAVGLKTVGWIMCKQEDAPDVIDTNGFCYCEGFRNPKVPKPK--- 246
DB 213 GNDVILFTTDCASEKM-LKGGTLQDLVATVD-----FGTGNNI-TQAFVQRKEPFGPL 265
QY 247 MTEWMTGWTKGPGIPQPAEDIAFVARFVQNNGSFFNYMYHGGTNF---GRTSS 302
DB 266 INSEFYGLDHDGKPHSTVKTATSLYNLLA-RGANVLYNFIGGTNFAYWNGANTP 324
QY 303 GLFIATSYDYDAPLDEYGLLEPKYHRLDLHKAIK-----LSEPALVSSVAAVTS 354
DB 325 YEQPTSYDYDAPLSEAGDLTK-KYFALREVQMFKEVPESGPIPPPTKAYGKVALRKF 383
QY 355 GSNQEAHVYKSKGACAAFLSNVDSRVKVTQNRPNYLPWVSISILPCKTAYVNTAQ 414
DB 384 KTVAAELGILCPNGP-----VKSPLYLFTTQVQYFGVLYRTLLPQDC----- 427
QY 415 VNSQSSIKMTAGGGLSWQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLWYTNV 474
DB 428 --SNPKFISPPNG-----VRDRA----- 445
QY 475 NIASNEGLKNGKDPYLVTVMSAGHLVHFVNGKLSGTGYTGLDNPKITYSGNVKLKRA 534
DB 446 -----YVSV-----DGVPQGIIDRLNMT-ALNIRKAGAT 474
QY 535 KISLLSVSGVLPNVG-----VHYDTWAGVLGVPVTLISGLNEGRNLAKQKWSKVG 585
DB 475 -LDILVENMGRVNGYGRINDFKGLISNMTINSTVLTNTVTFPLTEAM-VRNHLWG 532
QY 586 LKGESLSLHSLSGSSSVWVGRSLMAQKPLTWYKATFNAPGGNDPLALD-----MAS 641
DB 533 DEGH-LDGRSTNSDDL-----ILP-----TFYVGNFISPGIPDLPDQDTFIQ 580

QY 642 QGIWINGEGVGRHPW 656
DB 581 GQVWINGENLGRYPW 595
RESULT 10
BGAL_ASPNG
ID BGAL_ASPNG STANDARD; PRT; 1006 AA.
AC P29853;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase).
GN LACA.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=92273176; PubMed=1368193;
RA Kumar V., Ramakrishnan S., Teeri T.T., Knowles J.K., Hartley B.S.;
RT "Saccharomyces cerevisiae cells secreting an Aspergillus niger beta-
galactosidase grow on whey permeate.";
RL Biotechnology 10:82-85(1992).
RN [2]
RX SEQUENCE FROM N.A.
RC STRAIN-VTT-D-80144;
RA Hartley B.S., Ramakrishnan S., Kumar V.;
RT "DNA construct and modified yeast.";
RL Patent number WO9010703, 20-SEP-1990.
CC -1- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
GANGLIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCANS.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; L06037; AAA32696.1; -;
CC EMBL; S37150; AAC60538.1; -;
CC EMBL; A00968; CAA00105.1; -;
CC InterPro; IPR001944; Glyco_hydro_35.
CC Pfam; PF01301; Glyco_hydro_35; 1.
CC PRINTS; PR00742; GLHYDLASE35.
CC PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KW Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 1006 BETA-GALACTOSIDASE.
FT ACT_SITE 200 200 PROTON DONOR (POTENTIAL).
FT ACT_SITE 298 298 NUCLEOPHILE (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 422 422 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 739 739 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 777 777 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 805 805 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 206 206 S -> C (IN REF. 2).
SQ SEQUENCE 1006 AA; 109160 MW; 7157B28A83805488 CRC64;

Query Match		10.0%; Score 393; DB 1; Length 1006;
Best Local Similarity		23.1%; Pred. No. 1.2e-19;
Matches		198; Conservative 102; Mismatches 318; Indels 238; Gaps 38;
Qy	25	VSYDDRAIIINGKRIISGSIHPRSTP--QMPDIIQKADGGLDVIEYVFWNGHEP 82
Db	46	VTWDDKSLFINGERIMFSGBFH-PFRLPVKELQIDIPQVKALGFCVSYVDWALVEG 104
Qy	83	SPGVNFEGRYDLVRFIKWVRAGLYVNLGRIPYVCAEMNFGPPVWLKYVPGMEFRNN 142
Db	105	KPGEVRADGIDLEPFFDAEAGYLLARECPYINASSGGPFGWLQRYNG-TLRSSD 163
Qy	143	QPFVYAMQGFQKIVNMKSNLENFSQGPPIIMAIQIENEYGEVWEIGAPKAYTKWAAQ 202
Db	164	KAYLDATDNYVSHAATIAKYOI--TNGPPIILYQENETYSGCGVEFPDPVIMQYVED 221
Qy	203	MAVGLKTCVPHI-----MCKQEDADP-----VIDTCN-----GFYC-----EGF 237
Db	222	QARNAGVVIPLINDNASAGNNAFCTGKGAVDYGHDSYPLGFCANPTVWPMSGDLPTNF 281
Qy	238	R-----PNKPYKPKMTEVMTGWYTKFGGPIPORPAEDIAFVAR-FVONNGSF-----FNY 288
Db	282	RTLHLEQSPTTPYAIVERQGSYDPGWGPGFAACSELLNNEFERVYKNDFSQIATMNL 341
Qy	289	YMYHGGTNFGRTS--SGLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVS 346
Db	342	YMIFGGTWNGLYPNGY---TSYDYGSAVTESRNITREKYSSELKLLGNFAKVPGLYLA 398
Qy	347	SYAAVTSIG--SNOEAVHYRSKSGACAFL-----SNYDSRYSVKVTQNRNYPNLPWSIS 400
Db	399	SPGNLTTSYGAADTDLATVPLLGNSGTGFFVVRHSDYSSEESTSYKLR-----LPTSAGS 453
Qy	401	I-LPCKTAVNTAQVNSOSSSIKWTIP---AGGGL-----SWOSYNEETPTADSDT 448
Db	454	VTIQ-----LCGTILNGRDSKIHTVDHNVSGTNIIYSTAEVFTWKE-----ADGKVL 503
Qy	449	LTANGLME-----QKNVT-----461
Db	504	VLYGAGGHHELAISTKSNVTVIEGSEGISKQTSVVVVGWDVSTPRRIIQVGDILKIL 563
Qy	462	---RDSYLLWYMNVTIASNEGFLKCKDPYLTVMAGHVLHVFNVGKLSGTVYGILDN 518
Db	564	LLDRNSAYNVVWPOLATDGTSPGSTPEKVASIIVKAGYLVR-----TAY-----609
Qy	519	PKLYSGNVKLKRALINKISLSVSVGLPNVGVHYDTWAGVLPVTLISGLNRSRLAKQ 578
Db	610	--LKSG-LYLTADFNATTSVEV-IGVPSTAK-----NLFINGDKTSHYVDKNG-----654
Qy	579	KWSYKVLKGSLSLSLSG-----SSSVE---WVRGSLMAQKQPL-----616
Db	655	IWSATVDYNAPDISLPSLKLDWKYVDLPETQSSYDSDSLWPAADLKQTKNLTSLTPT 714
Qy	617	-----TWYKATFNAGNDPLALDM--ASMKGKQIWINGEGVGRHPGYIAQ 661
Db	715	SLYSSDYGFHTGYLLYRGHFTATGNSTFAIDTQGSAGFAGSVWLNGLYLG-SWTGLYAN 773
Qy	662	GDCSKCYAGTFNEKKCO-----TNGQBPQSQ 687
Db	774	SD-----YNATYNLPOLQAGKTYITVVDNMGLEENWTVGEDLMKSPRGISTSLPDGQ 828
Qy	688	RWYHVRSWLKPSGNL 703
Db	829	A---APISW-KLTGNL 840
RESULT 11		
ID	BGAM_HUMAN	STANDARD; PRT; 546 AA.
AC	P16279;	
DT	01-AUG-1990 (Rel. 15, Created)	
DT	01-AUG-1990 (Rel. 15, Last sequence update)	
DT	15-JUL-1999 (Rel. 38, Last annotation update)	
DE	Beta-galactosidase-related protein precursor.	

GN	GLB1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND PLACENTAL, PARTIAL SEQUENCE.	
RC	TISSUE=Testis;	
RX	MEDLINE=90062209; PubMed=2511208;	
RA	Morreau H., Gallart N.J., Gillemans N., Willemssen R.,	
RA	van der Horst G.T.J., D'Azzo A.;	
RT	"Alternative splicing of beta-galactosidase mRNA generates the	
RT	classic lysosomal enzyme and a beta-galactosidase-related protein.";	
RL	J. Biol. Chem. 264:20655-20663(1989).	
CC	-!- FUNCTION: THIS PROTEIN HAS NO CATALYTIC ACTIVITY.	
CC	-!- ALTERNATIVE PRODUCTS: BETA-GALACTOSIDASE AND BETA GALACTOSIDASE-	
CC	RELATED PROTEIN ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME	
CC	GENE.	
CC	-!- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M27508; AAA35599.1; --	
DR	PIR; B32688; B32688.	
DR	MIM; 230500; --	
DR	InterPro; IPR001944; Glyco_hydro_35.	
DR	Pfam; PF01301; Glyco_hydro_35; 1.	
KW	Signal; Alternative splicing; Glycoprotein.	
FT	SIGNAL 1 23	
FT	CHAIN 24 546 BETA-GALACTOSIDASE-RELATED PROTEIN.	
FT	CARBOHYD 26 26 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 367 367 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 411 411 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).	
FT	CARBOHYD 424 424 N-LINKED (GLCNAC. .) (POTENTIAL).	
SQ	SEQUENCE 546 AA; 60551 MW; 2B1A73EDAF9E966C CRC64;	
Query Match 5.8%; Score 227.5; DB 1; Length 546;		
Best Local Similarity 20.4%; Pred. No. 1.8e-08;		
Matches 150; Conservative 77; Mismatches 226; Indels 283; Gaps 29;		
QY	6	VLLLVICLLDFFSSVKAS-----VSYYDDRAIIINGKRIISGSIHPRSTPQMPD 59
DB	9	LLLLLVILLGLPTGLRNATORMFEIDYSRDSFLKDGQPFYISGSIHYSRVPFYWKDR 68
QY	60	IQAKADGGLDVITYVFWNGHEPSPGKYNFEGRYDLVRFIKWVRAGLYVNLGRIPYVCA 119
DB	69	LLKMKMAGLNAIQT-----82
QY	120	EWNFGGPPVWLKYVPGMEFRTNQPFKVMQGFVQKIVNMKSNLENFSQGGPIIMAIQIE 179
DB	83	-----LPG-----SCGVVGVGSPSAQ 97
QY	180	NEYGPVWEIGAPGKAYTKWAAQNAVLGKTVPTWIMCKQEDADPDPVIDTCNGFCYCEGRP 239
DB	98	DEASPL-----SEWRA-----SYNAGSNITD---AFISQ--RK 126
QY	240	NKPYKPKMTEVMTGWYTKFGGPIQRPADIAFVARFVQNGSGFFNYMYHGGTNF-- 297
DB	127	CEPKGPLINSEFYTGWLHDHWGPHSTIKTEAVASSLYDILA-RGASVNLVWFIGTNPAY 185
QY	298	--GRTSSGLFIATSYDYDAPLDEYGLLNEPKYGHLD-RHKAIKLSEPALVSS-----YA 349
DB	186	WNGANSPYAAQPTSYDYDAPLSEAGDLTE-KYFALRNIIQKFEKVPGEPIPPSTPKFAYG 244


```
Db 225 -----DD-----AYSDKSMKVTYAFNQFG-----PNCQRMPRARYG 257
QY 550 -VH-----YDTWAGVLG-----PVTSLNGSRNLAKQKWSYKVLKGSLSLHSLSGS 599
Db 258 LVHVANNNDPWTIATGSSNPILS-----EGSFTAPNE-SYK---KQVTIRIGCKTSS 310
QY 600 SSVEWV-----RGSLSMAQKQPLTWYKATFNAPGCGN 629
Db 311 SCSNWVQSTODVEFYNGAIFVSSGKYEGGNIYTKKE-----AFNVENG 354

RESULT 13
Y321_MYCGE
ID Y321_MYCGE STANDARD; PRT; 934 AA.
AC P47563;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical lipoprotein MG321 precursor.
GN MG321.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RA "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
(Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39713; AAC71543.1; -
CC TIGR; MG321; -
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 934 HYPOTHETICAL LIPOPROTEIN MG321.
FT LIPID 25 25 N-ACYL DIGLYCERIDE (PROTEINAL).
SQ SEQUENCE 934 AA; 102999 MW; EDAA036543736C6F1 CPG64;

Query Match 3.0%; Score 118.5; DB 1; Length 934;
Best Local Similarity 21.2%; Pred. No. 1.5;
Matches 149; Conservative 90; Mismatches 257; Indels 207; Gaps 38;

QY 135 GMEFRTNQPKVAMQGFQKIVNMKSENLFESQGGPIIMAIQIENEYGPVWEIGAPG 193
Db 185 GREVKQNNQPKVLSSKDFERGFEYIILSSNLGFRNRYFDLMLGLD-----VEKTVG--- 236
QY 194 KATYKAAQMAVGLKGVPMWIMQKQEDAPD-----PVDTGNGFCYCEGFR---PNK----- 241
Db 237 -----MDKNTGSSDGNKGKGIETDENVYRSDVNNKFNVL 275
QY 242 ----PYKPKMTEVWTGWTYKFGGPIQRPADIAF-----SVARFYQNG-----SFF 286
Db 276 TSPFPFLSMMSK-----EFTFPPIPTHKPKALKLGDSPKLYQNQNNRKILDQANT 328
```

```
QY 287 NY-YMYHGGTNFGRT---SSGLFIATSYD-----YDAPLDEYGLLNEPKYGHRLD 332
Db 329 NFDGIYGGVNAWKDTWSVGPIYYVESFNQAGIVFKRNIQIDAITP-----NLPK--TRQE 382
QY 333 LHKAIKLSEPALVS-----SYAAVTSLSNQBAHVYRSKSGACAAFLSNYDSRY- 381
Db 383 NEKPI-----PAIVSYFQPGATPEVYSSYIAGLSASAVPYSQOQDARSRENGTDLRLW 438
QY 382 -----SVKVTFFONRPNLPPWSISILPCKTAVYNTAOV-----NSOSSS 421
Db 439 KIQTAQSAQVTSYSGKY-----VANDSTVQLNANITETEAKFLYNSEEE 484
QY 422 IKMTPAG--GGLSQSYN-EETPTADSDTLT-----ANGLMEQKKNVTRDSSDYL 468
Db 485 ALTIRAGINGLINWKNLAIIDLPNSGDVNYSTVPFGIFKEKPANGTSGGTNDGIENDY- 543
QY 469 WYMTNVN-----IASNEGFLKNGKDPYLTVMSAGHV-LHVFVNGKLSGTVVGTLDNPKLT 522
Db 544 YYKINNQRGLIPEQTGTFQDKN-----VLDTATVKLSSYSTKTNGCAQVRTAST----- 595
QY 523 YSGNVKLKLAGINKISLSVSV-----GLPNVG-----VHYDTWAGVLGPVTLGLN 569
Db 596 -SGS-----SSQTSQVSSKQSVTKQSFISALKKVGFTGNPLHFN-----KLGNASLSSNQ 647
QY 570 EGSRLAKQKWSYKVLKGSLSLHSL-----SGSSSVWVRGSLMAQKOPLTWYKATFN 624
Db 648 VDIYNALKQALTELGNNGENLIPEIILGDAQGPTRENIWIG--LSVLGFSWSPDYD 705
QY 625 APGNDPLADMASMGKQIWIINGEGVGR-----HWPGYIAGDCSKCSYAGTFN--EKKC 678
Db 706 GVG---TWLDAATQ-----LNSEGIGEVITYNSGSHIVRTLLLAASQNNVFNQIENKL 755
QY 679 QTNCGQSQSRWYHVP RS--WLKPSGNLLVVFEEGNGPTGISL 719
Db 756 QNNTTTNGKMCWCSCITSADLKFDDPYVIKNGFTNGNGTSASL 798

RESULT 14
FLT3_HUMAN
ID FLT3_HUMAN STANDARD; PRT; 993 AA.
AC P36888; Q13414;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE FL cytokine receptor precursor (EC 2.7.1.112) (Tyrosine-protein kinase
DE receptor FLT3) (Stem cell tyrosine kinase 1) (STK-1) (CD135 antigen).
GN FLT3 OR STK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94119906; PubMed=7507245;
RA Small D., Levenstein M., Kim E., Carow C., Amin S., Rockwell P.,
RA Witte L., Burrow C., Ratajczak M.Z., Gewirtz A.M., Civin C.I.;
RA "STK-1, the human homolog of Flk-2/Flt-3, is selectively expressed in
RA CD34+ human bone marrow cells and is involved in the proliferation of
RA early progenitor/stem cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:459-463(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93357464; PubMed=8394751;
RA Rosnet O., Schiff C., Pebusque M.J., Marchetto S., Tonnelle C.,
RA Toiron Y., Birg F., Birnbaum D.;
RA "Human FLT3/FLK2 gene: cDNA cloning and expression in hematopoietic
RA cells.";
RL Blood 82:11110-11119(1993).
RN [3]
RP SEQUENCE OF 783-942 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=91169547; PubMed=2004790;
```


RT 1 in the t(8;13)(p11;q12) myeloproliferative syndrome.";

RL [3]

RN MEDLINE=99107818; PubMed=9889006;

RX MEDLINE=99107818; PubMed=9889006;

RA Kulkarni S., Reiter A.J., Smedley D., Goldman J.M., Cross N.C.P.;

RT "The genomic structure of ZNF198 and location of breakpoints in the

RT t(8;13) myeloproliferative syndrome.";

RL Genomics 55:118-121(1999).

[4]

RN SEQUENCE FROM N.A.

RA Blakey S., Wall M.;

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

[5]

RN SEQUENCE OF 152-1377 FROM N.A.

RX MEDLINE=98361795; PubMed=9694738;

RA Still I.H., Cowell J.K.;

RT "The t(8;13) atypical myeloproliferative disorder: further analysis of

RT the ZNF198 gene and lack of evidence for multiple genes disrupted on

RT chromosome 13.";

RL Blood 92:1456-1458(1998).

[6]

RN SEQUENCE OF 170-1020 FROM N.A.

RX MEDLINE=98167848; PubMed=9499416;

RA Smedley D., Hamoudi R., Clark J., Warren W., Abdul-Rauf M., Somers G.,

RT Venter D., Fagan K., Cooper C., Shipley J.;

RT "The t(8;13)(p11;q12) rearrangement associated with an atypical

RT myeloproliferative disorder fuses the fibroblast growth factor

RT receptor 1 gene to a novel gene RAMP.";

RL Hum. Mol. Genet. 7:637-642(1998).

[7]

RN SEQUENCE OF 621-1377 FROM N.A.

RX MEDLINE=98085877; PubMed=9425908;

RA Xiao S., Nalabolu S.R., Aster J.C., Ma J., Abruzzo L., Jaffe E.S.,

RT Stone R., Weissman S.M., Hudson T.J., Fletcher J.A.;

RT "FCFRL is fused with a novel zinc-finger gene, ZNF198, in the t(8;13)

RT leukaemia/lymphoma syndrome.";

RL Nat. Genet. 18:84-87(1998).

CC -!- FUNCTION: May function as a transcription factor.

CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).

CC -!- DISEASE: Involved in a t(8;13)(p12;q12) chromosomal translocation

CC which involves FCFRL AND ZNF198. The resulting transcript is a

CC possible candidate for stem cell leukemia lymphoma syndrome/SCLL.

CC -!- SIMILARITY: CONTAINS 5 MYM-TYPE ZINC FINGER.

CC -!- CAUTION: Ref.5 sequence differs from that shown due to a

CC frameshift in position 330.

CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts

CC in positions 330, 366, 1009 and 1017.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; Y13472; CAA73875.1; -

DR EMBL; AJ224901; CAA12204.1; -

DR EMBL; AJ007676; CAA07604.1; -

DR EMBL; AJ007677; CAA07604.1; JOINED.

DR EMBL; AJ007678; CAA07604.1; JOINED.

DR EMBL; AJ007679; CAA07604.1; JOINED.

DR EMBL; AJ007680; CAA07604.1; JOINED.

DR EMBL; AJ007681; CAA07604.1; JOINED.

DR EMBL; AJ007682; CAA07604.1; JOINED.

DR EMBL; AJ007683; CAA07604.1; JOINED.

DR EMBL; AJ007684; CAA07604.1; JOINED.

DR EMBL; AJ007685; CAA07604.1; JOINED.

DR EMBL; AJ007686; CAA07604.1; JOINED.

DR EMBL; AJ007687; CAA07604.1; JOINED.

DR EMBL; AJ007688; CAA07604.1; JOINED.

DR EMBL; AJ007689; CAA07604.1; JOINED.

DR EMBL; AJ007690; CAA07604.1; JOINED.

DR EMBL; AJ007691; CAA07604.1; JOINED.

DR EMBL; AJ007692; CAA07604.1; JOINED.

DR EMBL; AJ007693; CAA07604.1; JOINED.

DR EMBL; AJ007694; CAA07604.1; JOINED.

DR EMBL; AJ007695; CAA07604.1; JOINED.

DR EMBL; AJ007696; CAA07604.1; JOINED.

DR EMBL; AL137119; CAC42467.1; -

DR EMBL; AL138688; CAC16956.1; -

DR EMBL; AF060181; AAC23591.1; ALT_FRAME.

DR EMBL; AF035374; AAB88464.1; ALT_FRAME.

DR EMBL; AF012126; AAC01561.1; -

DR MIM; 602221; -

KW Transcription regulation; Nuclear protein; Chromosomal translocation;

Repeat; Zinc-finger.

FT ZN_FING 331 409 MYM-TYPE 1 (POTENTIAL).

FT ZN_FING 425 502 MYM-TYPE 2 (POTENTIAL).

FT ZN_FING 536 616 MYM-TYPE 3 (POTENTIAL).

FT ZN_FING 639 712 MYM-TYPE 4 (POTENTIAL).

FT ZN_FING 724 799 MYM-TYPE 5 (POTENTIAL).

FT SITE 913 914 BREAKPOINT FOR TRANSLOCATION TO FORM

FT ZNF198-EGFR1.

FT CONFLICT 388 416 VAOYDSSSEFCSTCLSLYEDKQNP -> GHLKWIQV

FT CONFLICT 657 659 NKV -> ASL (IN REF. 4).

FT CONFLICT 736 736 K -> G (IN REF. 1).

FT CONFLICT 766 767 MISSING (IN REF. 4).

FT CONFLICT 967 967 MISSING (IN REF. 6).

FT CONFLICT 1009 1010 DF -> IS (IN REF. 6).

FT CONFLICT 1016 1016 MISSING (IN REF. 6).

SQ SEQUENCE 1377 AA; 154910 MW; 2652D4C766492FF9 CRC64;

Query Match 2.8%; Score 110.5; DB 1; Length 1377;

Best Local Similarity 23.4%; Pred. No. 8.9;

Matches 93; Conservative 47; Mismatches 148; Indels 109; Gaps 22;

Qy 342 PALVSSAAVTSLSGNOEHAHYRSGACAAFLSNYDSRYSVKVTQN----- 389

Db 16 PVLGGSTAMATSL-----TNGNSFSGPANPLVRSNK-----FONSSVEDDDVVFI 63

Qy 390 RPYNLPPHSISILPCKTAVNTAQVNS-QSSSIKMTAGGSLWSQSYN-EETPTADSD 447

Db 64 EPVQPPSPVPVADQRTITTSKNEELQNDKSTPSSKELASOKGSVSETIVIDDEE 123

Qy 448 TLTANGLWEQ-----KNVTR--DSSDYLTWNTNVIASNEGFLKNGKDPYL- 491

Db 124 DMETNQGEKNSNFIERPPETKNTNDVDETSFSRSKVNAGMNGSGITTEPDSEIQ 183

Qy 492 ----TVMSAGHVLHVYVNGKLSGTVYGTLDNPKLITYSGNVKLR-----AGINKISLSVSV 543

Db 184 IANVTTLTETG-----VSSVNDGQLEN---TDGRDMNLMTHTVSLQNTNLGDVSN 230

Qy 544 GL--PNVGVDHTWNAVY-----LGPVTLGSLNEGRNLAKQKWSYKVLKGSLSLH 594

Db 231 GLQSNFGVNIQTPTPSLTQTKTGVP-----FNPGRMNVAGD-----VFQNGSATHH 280

Qy 595 SLSG--SSSWEVRGSLMAQKPLTWYKATFNAPGNDPLALDMSMGKGQIINGEGVG 652

Db 281 NPDWSISOSASEPRN-----QKQP-----GVDSLS-PVASLPK-QIFQPSVQQQ 322

Qy 653 RHWPGYIAQDCSKCSYAG-TFNEKK-----CQTNC 682

Db 323 PTKPVKVTCAKCKPLQKGTAYQRKSAHLFCSTTC 359

Search completed: June 24, 2002, 20:42:23

Job time: 230 sec


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:36:19 ; Search time 22.94 Seconds
(without alignments)
3032.636 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLVLLVLCLLDFSS.....VVFEWGGNPTGISLVRSR 724
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3924	100.0	724	2 T04340	beta-galactosidase
2	3043	77.5	731	2 T17002	probable beta-gala
3	2922.5	74.5	835	2 T06590	probable beta-gala
4	2863	73.0	832	2 S41889	beta-galactosidase
5	2838	72.3	729	2 T04269	probable beta-gala
6	2797	71.3	727	2 T47541	beta-galactosidase
7	2676	68.2	731	2 S16595	gene CARSR12 prote
8	2608	66.5	853	2 B85429	beta-galactosidase
9	2572	65.5	853	2 T04600	probable beta-gala
10	2259.5	57.6	839	2 C84685	probable beta-gala
11	2035.5	51.9	895	2 T00787	probable beta-gala
12	1909.5	48.7	828	2 S52393	beta-galactosidase
13	1647.5	42.0	780	2 D96803	probable beta-gala
14	1642	41.8	832	2 B84543	beta-galactosidase
15	1410.5	35.9	831	2 T05771	beta-galactosidase
16	1380.5	35.2	1036	2 T05687	beta-galactosidase
17	578	14.7	586	2 JC5618	beta-galactosidase
18	547	13.9	595	2 H95006	beta-galactosidase
19	543	13.8	595	2 C97879	beta-galactosidase
20	525.5	13.4	598	2 A57249	beta-galactosidase
21	518	13.2	612	2 B82756	beta-galactosidase
22	513.5	13.1	595	2 T29434	beta-galactosidase
23	495	12.6	677	2 A32611	beta-galactosidase
24	479.5	12.2	647	2 A37086	beta-galactosidase
25	429	10.9	616	2 C96755	Similar to acid be
26	413	10.5	746	2 T24978	hypothetical prote
27	393	10.0	982	2 A97210	beta galactosidase
28	365.5	10.0	1006	2 T31685	beta-galactosidase
29	365.5	9.3	681	2 T33381	hypothetical prote

30	269.5	6.9	778	2 B71164	probable beta-gala
31	265.5	6.8	787	2 C75068	probable beta-gala
32	227.5	5.8	546	2 B32688	beta-galactosidase
33	223	5.7	171	2 S37748	beta-galactosidase
34	146.5	3.7	1455	1 A48925	mannose receptor p
35	131	3.3	374	2 JC2124	major allergen Cry
36	129.5	3.3	1661	2 T31330	head-activator bin
37	128	3.3	2204	2 A70524	probable PPE prote
38	126.5	3.2	2902	2 C71953	toxin-like outer m
39	125	3.2	374	2 JC2123	major allergen Cry
40	120	3.1	1932	2 S53409	probable membrane
41	120	3.1	2154	2 P83088	hypothetical prote
42	119	3.0	686	2 AH0104	beta-galactosidase
43	118.5	3.0	934	2 E64235	hypothetical prote
44	118	3.0	1214	2 T30941	DNA polymerase - c
45	117.5	3.0	1449	2 T30552	glucosyltransferas

ALIGNMENTS

RESULT 1

T04340

beta-galactosidase (EC 3.2.1.23) II precursor - tomato

C:Species: Lycopersicon esculentum (tomato)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000

C:Accession: T04340

R:Smith, D.L.; Starrett, D.A.; Gross, K.C.

Plant Physiol. 117, 417-423, 1998

A:Title: A gene coding for tomato fruit beta-galactosidase II is expressed during fru

A:Reference number: Z15296; MUID:98289087

A:Accession: T04340

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-724 <SMT>

A:Cross-references: EMBL:AF020390; NID:g3299895; PIDN:AAC25984.1; PID:g3299896

A:Experimental source: strain Rutgers; tissue-type tomato fruit

C:Genetics:

A:Gene: Bgal4

C:Superfamily: beta-galactosidase bga

C:Keywords: glycosidase; hydrolase

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-724/Product: beta-galactosidase II #status predicted <MAT>

Query Match	100.0%	Score 3924	DB 2:	Length 724
Best Local Similarity	100.0%	Pred. No. 3.2e-265		
Matches 724	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Qy 1 MLRTNVLVLLVLCLLDFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPQMPDII 60

Db 1 MLRTNVLVLLVLCLLDFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPQMPDII 60

Qy 61 QKAKDGLDVIETIYVFWNGHEPSPKYNFEGHYDLVRFIKMVQRAGLYNLNIGIYVCAE 120

Db 61 QKAKDGLDVIETIYVFWNGHEPSPKYNFEGHYDLVRFIKMVQRAGLYNLNIGIYVCAE 120

Qy 121 WNFGGFPVWLKVPVGMFEFTNNOPEKVMQGFVQKIVNMKSENLFESOGGPIIMAIQEN 180

Db 121 WNFGGFPVWLKVPVGMFEFTNNOPEKVMQGFVQKIVNMKSENLFESOGGPIIMAIQEN 180

Qy 181 EYGPVEWEIGAPKAYTKWAAQMAVGLKTGVVWIMCKEDADPDVIDTCNGPYCGFRPN 240

Db 181 EYGPVEWEIGAPKAYTKWAAQMAVGLKTGVVWIMCKEDADPDVIDTCNGPYCGFRPN 240

Qy 241 KPYKPKMTTEWTKYHTEGFGPIQORPAEDIAFVSARFVQNNGSFFNYMHGGTFNGRT 300

Db 241 KPYKPKMTTEWTKYHTEGFGPIQORPAEDIAFVSARFVQNNGSFFNYMHGGTFNGRT 300

Qy 301 SSGLFATISYDADPLDEYGLLNEPKYGHRLDKAIKLSSEPALVSSVAAYTSLGSNOEA 360

Db 301 SSGLFATISYDADPLDEYGLLNEPKYGHRLDKAIKLSSEPALVSSVAAYTSLGSNOEA 360

```

QY 361 HYVRSKGACAAFLSNYDSRYSVKVTQFQNRPNLPPWISILPDCCKTAVYNTAQVNSQSS 420
|||||
Db 361 HYVRSKGACAAFLSNYDSRYSVKVTQFQNRPNLPPWISILPDCCKTAVYNTAQVNSQSS 420
QY 421 STKMTYPAGGLSWQSVNEETPTADSDTLTANGLWEQKNVTRDSDYLWYMTNVAIASNE 480
|||||
Db 421 STKMTYPAGGLSWQSVNEETPTADSDTLTANGLWEQKNVTRDSDYLWYMTNVAIASNE 480
QY 481 GFLKNGKDPYLTVMAGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVKLKAGINKISILLS 540
|||||
Db 481 GFLKNGKDPYLTVMAGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVKLKAGINKISILLS 540
QY 541 VSVGLPNVGVHYDTNAGVGLPVTLSGLNEGRNLAKQWSYKVKGLGESLSLHSLSGSS 600
|||||
Db 541 VSVGLPNVGVHYDTNAGVGLPVTLSGLNEGRNLAKQWSYKVKGLGESLSLHSLSGSS 600
QY 601 SVEWYRGLSMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIA 660
|||||
Db 601 SVEWYRGLSMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIA 660
QY 661 QGDCSKSYAGTFNEKKCOTNCGQPSQRYHVPRSWLKPSPGNLLVVFPEWGNPTGISLV 720
|||||
Db 661 QGDCSKSYAGTFNEKKCOTNCGQPSQRYHVPRSWLKPSPGNLLVVFPEWGNPTGISLV 720
QY 721 RRSR 724
|||||
Db 721 RRSR 724

RESULT 2
TI7002
probable beta-galactosidase (EC 3.2.1.23) precursor - apple tree
C:Species: Malus domestica (apple tree)
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C:Accession: TI7002
R:Ross, G.S.; Wegryn, T.; MacRae, E.A.; Redgwell, R.J.
Plant Physiol. 106, 521-528, 1994
A:Title: Apple beta-galactosidase. Activity against cell wall polysaccharides and charac
A:Reference number: 218645; MUID:95083752
A:Accession: TI7002
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-731 <ROS>
A:Cross-references: EMBL:L29451; NID:9507277; PIDN:AAAG62324.1; PID:9507278
A:Experimental source: strain Granny Smith; cortical
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase
F:1-23/domain: signal sequence #status predicted <SIG>
F:24-731/Product: probable beta-galactosidase #status predicted <MAT>

Query Match 77.5%; Score 3043; DB 2; Length 731;
Best Local Similarity 74.9%; Pred. No. 7.3e-204;
Matches 536; Conservative 96; Mismatches 80; Indels 4; Gaps 2;

QY 7 LLLLVICLLDFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDLIQAKDG 66
|||||
Db 11 ILLLFSCI--FSAASASVSDHKAIIINGKRIISGSIHYPRSTPQMPDLIQAKDG 67
QY 67 GLDVETVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLVNLRIGPYCAENWFGF 126
|||||
Db 68 GLDVITQTVFWNGHPSGPNYFEERYDLVRFKILVQOEGFLVNLRIGPYCAENWFGF 127
QY 127 PWLVKYVPGMEFRTNQPFKVMQGVOKIVNMKSENLFESQGPILMAQIENYGPVE 186
|||||
Db 128 PWLVKYVPGIAFRTDNEFPKAAQKFTKIVSMKAEKLFQFGPILISQIENYGPVE 187
QY 187 WEIGAPKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKYKP 246
|||||
Db 188 WEIGAPKAYTKWAAQMAVGLDTPVWIMCKQEDAPDPVIDTCNGFYCEFKPNKDYPK 247
QY 247 MWTEVMTGWTKFGGPIPORPAEDIAFSVARFVNQNSFFNYHYHGTNFGRTSSGLFI 306
|||||

```

```

Db 248 MWTEVMTGWTEFGGAVPTRPAEDVAFSVAREIQSGGSFLNYMYHGTNFGRTAGGPPM 307
QY 307 ATSYDYDAPLDDEYGLNNEPKYGHRLDLHKAIKLSEPALVSSVAAVTSLSGNSQEAHVYRSK 366
|||||
Db 308 ATSYDYDAPLDDEYGLPREPKWGHRLDLHKAIKSCESALVSDVPSVTKLGSNQEAEHVFKSE 367
|||||
QY 367 SGACAAFLSNYDSRYSVKVTQFQNRPNLPPWISILPDCCKTAVYNTAQVNSQSSIKMTP 426
|||||
Db 368 SD-CAAFLANVDKASVSVKFGGQYDLPPWISILPDCCKTEVNTAKVGSOSSQVQMT 426
|||||
QY 427 AGGGLSWQSVNEETPTADSDTLTANGLWEQKNVTRDSDYLWYMTNVAIASNEGFLNG 486
|||||
Db 427 VHSQFPMQSFIBETTSDETDTTLDGLYEQINIRDTTDLWYMTDITIGSDEAFLNG 486
QY 487 KDPYLTVMAGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVKLKAGINKISILLSVGLP 546
|||||
Db 487 KSPULTIFISAGHALNVFINGQLSGTVYGSLENPKLFSQNVNLRSGINKLALLSISVGLP 546
QY 547 NVGHYDTNAGVGLPVTLSGLNEGRNLAKQWSYKVKGLGESLSLHSLSGSSSVEWYR 606
|||||
Db 547 NVGTHFTWAGVGLPITLKLNSGTWDMGSKWTKYTKGLKEALGLHTVTGSSSVEWVE 606
QY 607 GSLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWINEGVGRHWPYIAQDCSK 666
|||||
Db 607 GPSMAEKQPLTWYKATFNAPGNDPLALDMSMGKQIWINESQSVGRHWPYIARGSCD 666
QY 667 CSYAGTFNEKKCOTNCGQPSQRYHVPRSWLKPSPGNLLVVFPEWGNPTGISLV 722
|||||
Db 667 CSYAGTYDDKKCRTHCGEPSQRYHVPRSWLTPGNTGNLLVVFPEWGGDPSRISLVER 722

RESULT 3
T06590
probable beta-galactosidase (EC 3.2.1.23) - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 19-May-2000
C:Accession: T06590
R:Carey, A.T.; Holt, K.; Picard, S.; Wilde, R.; Tucker, G.A.; Bird, C.R.; Schuch, W.;
Plant Physiol. 108, 1099-1107, 1995
A:Title: Tomato exo-(1-4)-beta-D-galactanase: isolation, changes during ripening in n
A:Reference number: 215780; MUID:95357407
A:Accession: T06590
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-835 <CAR>
A:Cross-references: EMBL:X83854; NID:971484; PIDN:CAA58734.1; PID:971485
A:Experimental source: cultivar Ailsa Craig; pericarp
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 74.5%; Score 2922.5; DB 2; Length 835;
Best Local Similarity 72.2%; Pred. No. 2.2e-195;
Matches 518; Conservative 94; Mismatches 102; Indels 3; Gaps 2;

QY 6 VLLLVICLLDFSSVKASVSDRAIIINGKRKILISGSIHYPRSTPQMPDLIQAKD 65
|||||
Db 7 MLLMLLCL--WVSGIASVSDHKAIIINGKRIISGSIHYPRSTPQMPDLIQAKE 64
QY 66 GLDVETVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLVNLRIGPYCAENWFGG 125
|||||
Db 65 GGDVDIQTIVFWNGHPSGKYNFEGRYDLVRFKMWQVAGLVNLRIGPYCAENWFGG 124
QY 126 FPVLKYVPGMEFRTNQPFKVMQGVOKIVNMKSENLFESQGPILMAQIENYGPV 185
|||||
Db 125 FPVLKYVPGISFRTNNEFPKAAQKFTKIVDMKAKLIETOGGPILISQIENYGPV 184
QY 186 EWEIGAPKAYTKWAAQMAVGLKTVGPWIMCKQEDAPDPVIDTCNGFYCEGFRPNKPKYKP 245
|||||
Db 185 EWEIGAPKAYTKWAAQMAVGLDTPVWIMCKQEDAPDPVIDTCNGFYCYDFTPNKANK 244
QY 246 MWTEVMTGWTKFGGPIPORPAEDIAFSVARFVNQNSFFNYHYHGTNFGRTSSGLFI 305
|||||

```

Db 245 KMTEAWTANTEFGGPPVDPYPAEDMAFAVARFIQTGGSFINYMYHGGTNGRSTGGPF 304
Qy 306 IATSYDYDAPLDEYGLLNEPKYGLRLDLHKAIKLSEPALVSSYAAVTSLGNOEAHVRS 365
Db 305 IATSYDYDAPLDEYGLLNEPKYGLRLDLHKAIKLSEPALVSSYAAVTSLGNOEAHVRS 365
Qy 366 KSGACAFLSNYSRYSYKVTFFONRPNLPPWSISILPDKCTAVYNTAQVNSOSSIKMT 425
Db 365 ESGACAFLSNYSRYSYKVTFFONRPNLPPWSISILPDKCTAVYNTAQVNSOSSIKMT 425
Qy 426 PAGGGLSWOSYNEETPTADSDTLTANGLWEOKNVTRDSSDYLVWYNTVNTASNEGFLN 485
Db 425 PYSRGFSHESFEDDAESHED-DFTFVGLLEDOINTROVSDYLVWYNTVNTASNEGFLN 485
Qy 486 GKDPYLTVMASGHLVHVFNKLSGTGYGTLDPNPKLTYSGNKLKLAGINKISLSSVSVGL 545
Db 484 GNPWLTVMASGHLVHVFNKLSGTGYGTLDPNPKLTYSGNKLKLAGINKISLSSVSVGL 545
Qy 546 PNVGPHYDTWAGVLPVTLTSLNEGRNLAKQKWSYKVLKGESLSLHSLSGSSVVEW 605
Db 544 PNVGPHYDTWAGVLPVTLTSLNEGRNLAKQKWSYKVLKGESLSLHSLSGSSVVEW 605
Qy 606 RGSLSMAQKOPLTWKATFNAPGNDPLALDMSMGKGOIWIINGEGVGRHWPYIAQGDPS 665
Db 604 EGSLSMAQKOPLTWKATFNAPGNDPLALDMSMGKGOIWIINGEGVGRHWPYIAQGDPS 665
Qy 666 KCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGNTLVVFEWGNPTGISLVRR 722
Db 664 VCNVTGWDFEKKCLTNCGEGSQRYHVPWSLKPNSGNTLVVFEWGNPTGISLVRR 722

RESULT 4
S41889
beta-galactosidase (EC 3.2.1.23) - garden asparagus
C:Species: Asparagus officinalis (garden asparagus)
C:Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 19-May-2000
C:Accession: S41889
R:King, G.A.; Davies, K.M.; Stewart, R.G.; Borst, W.M.
submitted to the EMBL Data Library, January 1994
A:Description: Identification and characterization of cDNA clones for asparagus mRNAs
A:Reference number: S41889
A:Accession: S41889
A:Molecule type: mRNA
A:Residues: 1-832 <KIN>
A:Cross-references: EMBL:X77319; NID:g452711; PIDN:CAA54525.1; PID:g452712
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 73.0%; Score 2863; DB 2; Length 832;
Best Local Similarity 71.4%; Pred. No. 3e-191;
Matches 514; Conservative 96; Mismatches 102; Indels 8; Gaps 5;

Qy 6 VLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKA 63
Db 6 VLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKA 63
Qy 64 KDGGLDVTETVYFVNGHEPSPGKYNFEGRYDLVRFKVMORAGLVNLRIGPYVCAEWNF 123
Db 64 KDGGLDVTETVYFVNGHEPSPGKYNFEGRYDLVRFKVMORAGLVNLRIGPYVCAEWNF 123
Qy 124 GGFVPLKYVPGMEFRTNPNQPFKVMQGVOKIVNMKSENLFESQGGPIIQAOTENYEG 183
Db 126 GGFVPLKYVPGMEFRTNPNQPFKVMQGVOKIVNMKSENLFESQGGPIIQAOTENYEG 183
Qy 184 PVEWIGAPGKAYTKWAQMAVGLKTGVPWIMCKOEDAPDPVIDTCNCFYCEGFRPNPKY 243
Db 186 PVEWIGAPGKAYTKWAQMAVGLKTGVPWIMCKOEDAPDPVIDTCNCFYCEGFRPNPKY 243
Qy 244 KPMWTEWGTWYTKFGGPIQORPAEDIAFSVARFVQNGSFFNYMYHGGTNGRSTSSG 303
Db 246 KPMWTEWGTWYTKFGGPIQORPAEDIAFSVARFVQNGSFFNYMYHGGTNGRSTSSG 303

Qy 304 LFIATSYDYDAPLDEYGLLNEPKYGLRLDLHKAIKLSEPALVSSYAAVTSLGNOEAHVY 363
Db 306 PFISTSYDYDAPLDEYGLLNEPKYGLRLDLHKAIKLSEPALVSSYAAVTSLGNOEAHVY 363
Qy 364 RSKSACAFLSNYSRYSYKVTFFONRPNLPPWSISILPDKCTAVYNTAQVNSOSSIK 423
Db 366 RSKSACAFLSNYSRYSYKVTFFONRPNLPPWSISILPDKCTAVYNTAQVNTATMK 424
Qy 424 MTPAGGGLSWOSYNEETPTADSDTLTANGLWEOKNVTRDSSDYLVWYNTVNTASNEGFL 483
Db 425 MTPAGGGLSWOSYNEETPTADSDTLTANGLWEOKNVTRDSSDYLVWYNTVNTASNEGFL 483
Qy 484 KNGKDPYLTVMASGHLVHVFNKLSGTGYGTLDPNPKLTYSGNKLKLAGINKISLSSVSV 543
Db 483 KNGKDPYLTVMASGHLVHVFNKLSGTGYGTLDPNPKLTYSGNKLKLAGINKISLSSVSV 543
Qy 544 GLPNVGNHFTWNTGVLGPTVLTGLNEGRDLNLSQKWTYQIGLHGETLSLHSLTSSNVE 602
Db 543 GLPNVGNHFTWNTGVLGPTVLTGLNEGRDLNLSQKWTYQIGLHGETLSLHSLTSSNVE 602
Qy 604 WYRGSLSMAQKOPLTWKATFNAPGNDPLALDMSMGKGOIWIINGEGVGRHWPYIAQGD 663
Db 603 WYRGSLSMAQKOPLTWKATFNAPGNDPLALDMSMGKGOIWIINGEGVGRHWPYIAQGD 663
Qy 664 CSKCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGNTLVVFEWGNPTGISLVRRS 723
Db 660 CSKCSYAGTFNEKKCOTNCGQPSQRYHVPWSLKPNSGNTLVVFEWGNPTGISLVRRS 719

RESULT 5
T04269
probable beta-galactosidase (EC 3.2.1.23) - Arabidopsis thaliana
C:Alternate names: protein F20B18.250
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 19-May-2000
C:Accession: T04269
R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 215263
A:Accession: T04269
A:Molecule type: DNA
A:Residues: 1-729 <BEV>
A:Cross-references: EMBL:AL049483
A:Experimental source: cultivar Columbia; BAC clone F20B18
C:Genetics:
A:Map position: 4
A:Introns: 58/3; 90/3; 128/2; 150/3; 181/3; 229/3; 259/2; 294/3; 323/1; 362/3; 416/3;
A:Note: F20B18.250
C:Superfamily: beta-galactosidase bga
C:Keywords: glycosidase; hydrolase

Query Match 72.3%; Score 2838; DB 2; Length 729;
Best Local Similarity 69.7%; Pred. No. 1.4e-189;
Matches 502; Conservative 104; Mismatches 108; Indels 6; Gaps 5;

Qy 7 LLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKADG 66
Db 11 LLLLVICLLDFFS--SVKASVYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLOKADG 66
Qy 67 GLDVIETVYFVNGHEPSPGKYNFEGRYDLVRFKVMORAGLVNLRIGPYVCAEWNFGF 126
Db 71 GLDVIETVYFVNGHEPSPGKYNFEGRYDLVRFKVMORAGLVNLRIGPYVCAEWNFGF 130
Qy 127 PWLKYVPGMEFRTNPNQPFKVMQGVOKIVNMKSENLFESQGGPIIQAOTENYEGPIE 186
Db 131 PWLKYVPGMEFRTNPNQPFKVMQGVOKIVNMKSENLFESQGGPIIQAOTENYEGPIE 190
Qy 187 WEIGAPGKAYTKWAQMAVGLKTGVPWIMCKOEDAPDPVIDTCNCFYCEGFRPNPKYK 246
Db 191 WEIGAPGKAYTKWAQMAVGLKTGVPWIMCKOEDAPDPVIDTCNCFYCEGFRPNPKYK 250
Qy 247 MWTEWGTWYTKFGGPIQORPAEDIAFSVARFVQNGSFFNYMYHGGTNGRSTSSGLFI 306

QY 62 KAKDGLDVIETVFWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAEW 121
Db 70 KAKDGDIDVIETVFWNLHEPSPGKYDFEGRNDLRFVFKTIHKAGLYAHLRIQPVYCAEW 129
QY 122 NFGGPPVWLKYYPGMEFRNNOFPKVMQGFQKIVNMKSENLPESQGGPIIQAQIENE 181
Db 130 NFGGPPVWLKYYPGISFRDNEPFRKMGFTFRIELMKSENLPESQGGPIIILSQIENE 189
QY 182 YGPVEWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKOEADAPDPVIDTCNGFYCEGFRPNK 241
Db 190 YGRQGLIGAEGHNTWTAAKMAIATEGVPWIMCKEADAPDPVINTCNGFYCDSEFAPNK 249
QY 242 YKPKMTEVWTGWTYKFGGPIQORPAEDIAFSAVFQVNGNSFFNYMYHGGTNGFRGTS 301
Db 250 YKPKLIWTEAWSGWTEFGGPMHRPQDLAFGVARFQKGGSFVNYMYHGGTNGFRGTA 309
QY 302 SGLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOE-- 359
Db 310 GGFVETTSYDYDAPIDEYGLIRQPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOE-- 369
QY 360 -----AHVYRSKGACAAFLSNYSRYKVTQFQNRPNLPWISILPDCCKTAVNTA 413
Db 370 IYERFAHYSAESGCSAFLANYDTESARVLFNNVHYNLPWISILPDCRNAVNTA 429
QY 414 QVNSOSSIKMTPAGGGLSWQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLYMYTN 473
Db 430 KYSN-----FQWESYLEDSLDDSDSTFTTHGLLEQINVTNRTSDYLWYMTS 476
QY 474 VNIASEGLFKGKDPYLTVMSAGHVLHVFNGLSGVYGTGLDNPKLITYSNVLRAGI 533
Db 477 VDIGSESLHGGEPLTLIQSTGHAVHIFVNGQLSGAFGRNRRFTYQCKINLHSGT 536
QY 534 NKISLLSVSVGLPNVGVHYDTWNAAGVLPVTLNGLNEGRNLAKQKWSYKVLKGSLSL 593
Db 537 NRIALLSVAVGLPNVGGHESWNTGILGPVALHGLSQKMDLSQKWTYQVGLKGEANWL 596
QY 594 HSLSGSSVWVYRGLSMAQK-OPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGV 652
Db 597 APPTNTPSIGWMDASILTQKQPPLTWHTYFDAPEGNEPLALDMEGCKQIWIINGESIG 656
QY 653 RHWPGYIAQDCSKSYAGTFNEKKCQTCNGQPSQRWYHVPWSLKPNSGILLVFEWGG 712
Db 657 RYWTAF-ATGDCSHCSYTGTYKPNKQTCGCGQPTQRWYHVPRAWLKPSONLLVIFEELGG 715
QY 713 NPTGSLVRRS 723
Db 716 NPSTVSLVRRS 726
RESULT 10
C84685
probable beta-galactosidase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: C84685
R.; Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-769, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84685
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <STO>
A:Cross-references: GB:AE002093; NID:g4510395; PIDN:AAD21482.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28470
A:Map position: 2
C:Superfamily: beta-galactosidase bga

Query Match 57.6%; Score 2259.5; DB 2; Length 839;
Best Local Similarity 57.1%; Pred. No. 3.2e-149;
Matches 418; Conservative 111; Mismatches 176; Indels 27; Gaps 10;
QY 3 RINVLILLVICLLDFESSVKASVSYDDRAIINGRKILISGSIHYPRSTPOMWPDLOK 62
Db 6 KHEMILLILIV--VAATAANVTYDHRALVDGRRKVLISGSIHYPRSTPEMPPELOK 63
QY 63 AKDGLDVIETVFWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAEW 122
Db 64 SKDGLDVIETVFWNGHEPSPGKYNFEGRYDLVRFIKMVORAGLYVNLRIQPVYCAEW 123
QY 123 FGGFPVWLKYYPGMEFRNNOFPKVMQGFQKIVNMKSENLPESQGGPIIQAQIENEY 182
Db 124 YGGFPVWLHFPVGIKFRDNEPFRKMGFTFRIELMKSENLPESQGGPIIILSQIENEY 183
QY 193 GPEWEIGAPKAYTKWAAQMAVGLKTGVPWIMCKOEADAPDPVIDTCNGFYCEGFRPNK 242
Db 184 GNIDSAYGAAKASYIKWSASMAISLDTGVPWIMCQOTDAPDPMINTCNGFYCDQFTPSN 243
QY 243 YKPKMTEVWTGWTYKFGGPIQORPAEDIAFSAVFQVNGNSFFNYMYHGGTNGFRGTS 302
Db 244 NKPKMTEWNSGFWLFGGDPSPYRPVEDLAFVAFYQGGTFQNYMYHGGTNGFRDTS 303
QY 303 GLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSSYAAVTSLSGNOEAVH 362
Db 304 GPLISTSYDYDAPIDEYGLLRQPKWHLRLHKAIKLCEDEALIAITDPTTISLGSLEAAV 363
QY 363 YRSKSGACAAFLSNYSRYKVTQFQNRPNLPWISILPDCCKTAVNTAQVNSQSSI 422
Db 364 YTESGSCAAFLIANVDTKSDATVTENGKSYNLPWISILPDCCKNVAENTAKV--KFNSI 421
QY 423 KMTAGG-----GLSQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLYMYTNVNTA 477
Db 422 SKTPDGGSSAELGSQW-SYIKPIGSKADAFKPLGLEQINTTADKSDLYSLRTDIK 480
QY 478 SNEGFLKNGKDPYLTVMSAGHVLHVFNGLSGVYGTGLDNPKLITYSNVLRAGI 533
Db 481 GDETFDEGSKAVLHIESILGQVYAFINGKLKAGSGHKGKISLDIP-----INLVGT 533
QY 534 NKISLLSVSVGLPNVGVHYDTWNAAGVLPVTLNGLNEGR-NLAKQKWSYKVLKGSLS 592
Db 534 NTFIDLLSVTVGLANGAFDFLVGAGITGPVTLKSAKSSIDLSAQQWTYQVGLKGEDTG 593
QY 593 LHLSLGSSSVWVYRGLSMAQK-OPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGV 652
Db 594 LATVDSS---EWVSKSPLTKQPLWYKTFDAPSGSEPAIDFTGCKGIATWVNGQSIG 650
QY 653 RHWPGYIA-QGDCSK-CSYAGTFNEKKCQTCNGQPSQRWYHVPWSLKPNSGILLVFEW 710
Db 651 RYWTPTSIAGNGGCTESCDYRGSYRANKCLKNGKPSQTLYHVPWSLKPNSGILLVFEW 710
QY 711 GSNPTGISLVR 722
Db 711 GGDPTQISFATK 722
RESULT 11
T00787
probable beta-galactosidase (EC 3.2.1.23) F24L7.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00787; G84737
R.; Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, February 1998
A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.
A:Reference number: Z14204
A:Accession: T00787
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-895 <ROU>
A:Cross-references: EMBL:AC003974; NID:g2914688; PID:g2914710
A:Experimental source: cultivar Columbia

QY 480 EGFLLKNGKDPYLTVMSAGHVLHVFVNGKLSGTVYGTLDNPKLTYSGNVNKLRAQINKISLL 539
Db 485 DLSKKGGKGNLRIASLGHALHVLNGEYLGNGHSGHEEKSFVQKDPVTLKEGENHLTML 544
QY 540 SVSGLPNVGVHYDTWAGVLGP--VTLGSLNGESRLAQ-KWSYKVLKLGESLSLHSL 596
Db 545 GVLTGFPDSSGYMB---HRYTGRPSVSILGLSGTDLDTTEENKGNKVGMEGERLGHAE 601
QY 597 SGSSSVWVRGSLMAQKOP-LTWYKATFNAPGGNDPLALDMASMGKQIWIINGEGVGRHW 655
Db 602 EGLKKVWERAS---GREPCMTWQTYFDAPESQSAAIRMNGMGKGLIWNNGEGVGRYW 658
QY 656 PGVIAQDCSKSVAGTFNEKKQTCNGQPSQRWYHVPWSLWPSGNLLVVFFEE 709
Db 659 MSFLSP-----LGQPTQIEYHIPRSLPKKLLLVIFEE 692

RESULT 15
T05771
beta-galactosidase homolog M4E13.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 02-Sep-2000
C:Accession: T05771
R:Bevan, M.; Purnelle, B.; Boutry, M.; Coffeau, A.; Hoheisel, J.; Mewes, H.W.; Mayer, K.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15451
A:Accession: T05771
A:Molecule type: DNA
A:Residues: 1-831 <BEV>
A:Cross-references: EMBL:AL022023
A:Experimental source: cultivar Columbia; BAC clone M4E13
C:Genetics:
A:Map position: 4
A:Introns: 26/1; 83/3; 115/3; 153/2; 169/3; 217/3; 249/2; 284/3; 352/3; 408/3; 467/2; 500/3
A:Note: M4E13.70
C:Superfamily: beta-galactosidase bga

Query Match 35.9%; Score 1410.5; DB 2; Length 831;
Best Local Similarity 39.8%; Pred. No. 4.2e-90;
Matches 294; Conservative 117; Mismatches 232; Indels 95; Gaps 16;

QY 2 LRNVLLLVICLLDFSSVKA-----SVSYDDRAIIINGKRKILISGSIHP 49
Db 19 LRKPNLTVLVLLSSSSFAAKDKKKSNKEVYDGTSLIDGKRELLYSGSIHP 78
QY 50 RSTPQWPDLIQAKDGLDVIETYPWNGHEPSPGKYNFEGYDLYRFFIKWQRAGLYV 109
Db 79 RSTPEMWPSTIKRAQOGGLTIQTYFVWVHEPQQGKFNFSGRADLVKFKLIKQNGMY 138
QY 110 NLRIGPVVCAEWNFSGFPVWLKYVPGMEFTNNQPPKVMQGFVKIVNMKSENLFESQ 169
Db 139 TRLGPFIOAEWHG-----YITRDHN-----TAGAYRK----- 169
QY 170 GPTIIMAEIENXGPVWEIGAPGKAYTKWAAQMAVGLKTGPWIMCKQEDAPDPVIDTC 229
Db 170 -----IENXSAVQRAYKQDGLNFKWASNLVDSMKLGPWCMCKQNDAPDPMINAC 221
QY 230 NGFYC-BGF-RPNKYPKPKWTEVWGTGWYTKFGGPIQORPAEDIAFSVARFVNNGSFFN 287
Db 222 NGRHCGDTFFGPNRENKPSLWNTWTQFRVFGDPPTQRSVEDIAYSVARFFSKNGTHVN 281
QY 288 YMYVHGNTNFRGTSGLFIATSYDYDAPLDEYGLLNEPKYGHLDLHKAIKLSEPALVSS 347
Db 282 YMYTHGNTNFRGRTSAH-YVTRYDDAPLDEYGLEKEPKYGHLDLHNLNMLCKPLWNG 340
QY 348 YAAVTSLSGNOEAHVYRSK-SGACAAFLSNYDSRYSVKVTQFNRPYNLPWPSISILPDC 406
Db 341 QPKTEKPGKDEIRYEQPGTKTCAAFANNTEAAETIKFGREYVIAPRISILPDC 400
QY 407 TAVYNTAQVNSQSSS----IKMTPAGGSLWSQSYNEETPTADDDSTLTANGLWQKNVTRD 463
Db 401 TVVYNTAQIVSQHTSRNFMKSKRANKKDFKFTETLPSKLEGNYSIPVELY---GLTKD 457

QY 464 SSDYLWMTNVTNVIASNEGFLKNGKDPYLTVMSAGHVLHVFVNGKLSGTVYGTLDNPKLTY 523
Db 458 KTDYGWYTTSFKVHKHNLPTKKGVKTFVRIASLGHALHAWLNGEYLGSGHSGHEKSFVF 517
QY 524 SGNVKLKRAKINKISLLSVSVGLPNVGVHYDTWAGVLGPVTLGSLNGESRLMAK-QKWSY 582
Db 518 QKVTLKAGENHLVMLGLVLTGFPDSSGYMBHRYTGPQG-LSILGTSGLDLETSSKWN 576
QY 583 KVGLKGESLSLHSLSGSSSVWVRGSLMAQKOP-LTWY-----KATFNAPGGNDP 631
Db 577 KIGMEGKLGILHTEGLKKVEWKK---FTGKAPGLTWYQKFSKCEETLQTYFDAPESVSA 633
QY 632 LALDMASMGKQIWIINGEGVGRHWPGYIAQDCSKSYAGTFNEKKQTCNGQPSQRWYH 691
Db 634 ATIRMHGCKGLIWNNGEGVGRYMQSFLSP-----LGQPTQIEYH 673
QY 692 VPSWLKPSGNLLVVFFEE 709
Db 674 IPRSLKPKKNLLLVIFEE 691

Search completed: June 24, 2002, 20:38:50
Job time: 151 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:33:54 ; Search time 14.86 Seconds
(without alignments)
1190.049 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLILLVLCILDFSS.....VVFEEGNGNPTGISLVRKR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 segs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2920	74.4	838	2	US-08-696-944-19
2	2896	73.8	730	2	US-08-696-944-2
3	2676	68.2	731	2	US-08-696-944-20
4	485	12.4	668	4	US-09-436-605-2
5	396.5	10.1	911	1	US-08-596-985-2
6	395	10.1	1007	2	US-08-551-459-4
7	146.5	3.7	1455	3	US-08-840-062-5
8	141	3.6	30	2	US-08-687-372-1
9	137	3.5	30	2	US-08-687-372-7
10	125	3.2	374	3	US-08-467-023-2
11	118	3.0	993	1	US-07-977-451-4
12	118	3.0	993	1	US-08-252-517-4
13	118	3.0	993	1	US-07-906-397A-4
14	118	3.0	993	1	US-08-222-299-4
15	118	3.0	993	1	US-08-601-891-4
16	118	3.0	993	2	US-09-021-324-4
17	118	3.0	993	2	US-08-434-878-4
18	118	3.0	993	5	PCT-US92-09893-4
19	118	3.0	993	5	PCT-US95-03718-4
20	118	3.0	1160	5	PCT-US92-05401-4
21	113	2.9	518	3	US-09-043-123-2
22	110	2.8	30	2	US-08-687-372-6
23	109	2.8	367	3	US-08-467-023-95
24	108	2.8	993	1	US-08-183-211-2
25	108	2.8	993	5	PCT-US95-00176A-2
26	107	2.7	566	2	US-07-862-588B-4
27	107	2.7	920	3	US-08-930-996A-8

Sequence 17, Appl
Sequence 17, Appl
Sequence 17, Appl
Patent No. 5268270
Sequence 48, Appl
Sequence 9, Appl
Sequence 9, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 70, Appl
Sequence 15, Appl
Sequence 9, Appl
Sequence 28, Appl
Sequence 11, Appl

28 107 2.7 1248 2 US-08-348-353-17
29 107 2.7 1248 2 US-08-465-965-17
30 107 2.7 1248 3 US-08-465-966-17
31 106.5 2.7 1507 6 5268270-2
32 105.5 2.7 2048 4 US-09-268-347-48
33 105 2.7 1338 2 US-08-728-470-9
34 105 2.7 1338 4 US-08-719-641-9
35 105 2.7 1599 2 US-08-617-697-9
36 104.5 2.7 654 1 US-08-392-828C-2
37 104.5 2.7 654 3 US-09-330-945-2
38 103.5 2.6 1702 4 US-08-296-791-5
39 103.5 2.6 1702 5 PCT-US95-10661A-5
40 103 2.6 686 3 US-08-947-965-70
41 103 2.6 1222 2 US-08-682-517-15
42 103 2.6 1252 2 US-08-682-517-9
43 102.5 2.6 1104 4 US-09-268-347-28
44 102.5 2.6 1104 4 US-09-268-347-34
45 101.5 2.6 591 4 US-09-377-155-11

ALIGNMENTS

RESULT 1
US-08-696-944-19
; Sequence 19, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 838 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-944-19

Query Match 74.4%; Score 2920; DB 2; Length 838;
Best Local Similarity 71.6%; Pred. No. 6.6e-254;
Matches 517; Conservative 96; Mismatches 105; Indels 4; Gaps 2;
Qy 1 MLRTNVLILLVLCILDFSSVKASVSDYDRAIIINGKRLISGSIHYPRSTPQWPDLI 60
Db 6 ILMLNVLVL---LGSVFTASVSDYDRAIIINGKRLISGSIHYPRSTPQWPDII 62
Qy 61 QKAKDGLDVIETVFWNGHEPSPGKYNFEGRYDLVRFKRWQAGLYVNLRIGPYVCAE 120

Db 63 QKAKGGVDVDTQYVFWNGHEPQQKYFEGRYDLVKIKLVHQAGLVXHLRVGPYACAE 122
Qy 121 WNFEGFPVWLKYNVPCMEFTNPNPKVAMOGFVOKIVNMKSENLFESQGGPIITMAQLEN 180
Db 123 WNFEGFPVWLKYNVPCISRTDNGPEKAAQKQFATAKIVNMKAERDIETQGGPIILQSLEN 182
Qy 181 EYGVPEWIEGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDAPDPVIDTCNGFYCEGFRPN 240
Db 183 EYGPMEWELGAPGKSYAOWAAKMAVGLDTGVPMWCKQDDAPDPPIINACNGFYCDYFSPN 242
Qy 241 KPYKPKMTWETWGTWYKFGGPIQPPAEDIAFSVARVQNNQSFNFYMYHGGTNFGRT 300
Db 243 KAYKPKIWEATAWTFGTGPNVPYPAEDLAFSAVAKIQRGGSFNYMYHGGTNFGRT 302
Qy 301 SSGFLIATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAATSLGNSQEA 360
Db 303 AGGPFIAATSYDYDAPLDEYGLLRQPKWHLKDLHRAIKLCEPALVSGDPVATLGHQOEA 362
Qy 361 HVIYRSKAGCAAFLSNYSRYSVKVTQFNRPNLPPWISISILPCKTAVYNTAQVNSQSS 420
Db 363 HVIYRSKAGCAAFLANYPDQHSFATVSPANRHNLPWISISILPCKNTVNTARIGQSA 422
Qy 421 SIKWTPAGGGLSWOSYNEETPTADSDTLTANGWEQKNVTRDSDYLWYNTVNNIASNE 480
Db 423 QMKMTVPVSRGLPWOSNEETSSYEDS-SFTVVGLEQINTTRDVSVDYLWYSTVDKIDRE 481
Qy 481 GFLKNGKDPYLTVMASAGHLVHFVNGKLSGVYGTLDNPKLTSYGNKLRAGINKISLLS 540
Db 482 KFLRGKKWPLTIVMSAGHALHFVNGQAGTAYGSLERPKLTFKAVNLRAGVKNKISLLS 541
Qy 541 VSVGLPNVGHYDTWNAAGVLGPTVLSGLNEGRNLAKQKWSYKVLKGESLSLSLSSGSS 600
Db 542 IAVGLPNIGPHFETWNAAGVLGPTVLSGLDEGRDLTWQKWFYKVLKGESLSLSLSSGP 601
Qy 601 SVEWVRSGLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGVGRHWPGVIA 660
Db 602 SVEWVRSGLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGVGRHWPGVIA 661
Qy 661 QGDCSKSYAGTFNEKKCOTKCGPQSRWYHVPKSWLKPNSNLLVVFPEWGNPTGSLV 720
Db 662 SGCSVCNCTGWFDEKKCLTNCGEGSQRWYHVPKSWLKPNSNLLVVFPEWGNPTGSLV 721
Qy 721 RR 722
Db 722 KR 723

RESULT 2

us-08-696-944-2
; Sequence 2, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 730 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-696-944-2

Query Match 73.8%; Score 2896; DB 2; Length 730;
Best Local Similarity 71.4%; Pred. No. 7.6e-252;
Matches 520; Conservative 94; Mismatches 100; Indels 14; Gaps 5;
Qy 1 MLRTN--VLLLV--ICLLDFSSVKASVSDRAIINGKRKILISGSIHYPRSTPQM 55
Db 12 MSRRNFHMLVLLLFVWCY-----VTASVTDHKAIMINGQRRLISGSIHYPRSTPQM 65
Qy 56 WPDLIQAKAGGLDVIETVFWNGHEPSPGKYNEGRYDLVRFKIMVORAGLYVNLRIQP 115
Db 66 WPDLIQAKAGGLDVIETVFWNGHEPSPGKYNEGRYDLVRFKIMVORAGLYVNLRIQP 125
Qy 116 YVCAEWNFGGPPVWLKYPVGMFEFTNNOPFKVAMOGFVOKIVNMKSENLFESQGGPIIM 175
Db 126 FICAEWFGGPPVWLKYPVGMFEFTNNOPFKVAMOGFVOKIVNMKSENLFESQGGPIIL 185
Qy 176 AQIENEGVPWEWELGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDAPDPVIDTCNGFYCE 235
Db 186 SQIENEGVPWEWELGAPGKAYTKWAQAQMAVGLTKGVPMWCKQEDALDPIIDTCNGFYCE 245
Qy 236 GFRPNKYPKPMWTEVMTWYTKFGGPIQPPAEDIAFSVARVQNNQSFNFYMYHGGT 295
Db 246 NFTPKNYKPKLWTEWMTWYTKFGGPIQPPAEDIAFSVARVQNNQSFNFYMYHGGT 305
Qy 296 NFGRTSSGLFTATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAATSLG 355
Db 306 NFGRTSSGLFTATSYDYDAPLDEYGLNRPKYGHLRDLHKAIKLSEPALVSSYAATSLG 365
Qy 356 SNOEAHVYRSKSGACAAFLSNYSRYSVKVTQFNRPNLPPWISISILPCKTAVYNTAQV 415
Db 366 KNLEVLHYKTES-ACAAFLANYNTDYSTQVKEGNGQYDLPWISISILPCKTEVENTAKV 424
Qy 416 NSQSSSIKMTPAGGGLSWQSYNEETPTADSDTLTANGWEQKNVTRDSDYLWYNTVWN 475
Db 425 NSPRLHRRMTVPVNSAFAMQSYNEEPASSSENDDPVTGYALWBEQVGVTRDSDYLWYNTVWN 484
Qy 476 IASNEGFLKNGKDPYLTVMASAGHLVHFVNGKLSGVYGTLDNPKLTSYGNVKLKLAGINK 535
Db 485 IGPND--IKDGKWPVLTAWSAGHLVNFVINGQYATAGTAYGSLDDPRLTFQSQVNLRVGNKK 542
Qy 536 ISLLSVSVGLPNVGVHYDTWNAAGVLGPTVLSGLNEGRNLAKQKWSYKVLKGESLSLSHS 595
Db 543 ISLLSVSVGLANVGTHTETWNTVGLVPTLTGLSSGTWDLKQKWSYKVLKGESLSLSHT 602
Qy 596 LSGSSSVWVRSGLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGVGRHW 655
Db 603 EAGSNSVEWVRSGLMAQKQPLTWYKATFNAPGNDPLALDMSMGKQIWIINGEGVGRHW 662
Qy 656 PGYTAQGDCKSKSYAGTFNEKKCOTKCGPQSRWYHVPKSWLKPNSNLLVVFPEWGNPT 715
Db 663 PGKARGNCGNCTAGTYTDTKCLANCGPQSRWYHVPKSWLKPNSNLLVVFPEWGNPT 722
Qy 716 GISLVRS 723
Db 723 GIALVERT 730

RESULT 3
US-08-696-944-20
; Sequence 20, Application US/08696944
; Patent No. 5981831
; GENERAL INFORMATION:
; APPLICANT: Sumant CHENGAPPA
; APPLICANT: Susan A. HELLYER
; APPLICANT: John S. REID
; APPLICANT: Jacqueline DE SILVA
; TITLE OF INVENTION: No. 5981831el Exo-(1-4)-Beta-D Galactanase
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,944
; FILING DATE: 23-AUG-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/00372
; FILING DATE: 23-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9403423.8
; FILING DATE: 23-FEB-1994
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-944-20

Query Match 68.2%; Score 2676; DB 2; Length 731;
Best Local Similarity 67.6%; Pred. No. 4.8e-232;
Matches 488; Conservative 89; Mismatches 137; Indels 8; Gaps 5;
Qy 6 VLLLLVICLLDFESSVKASVYDDRAIIINGKRKTLISGSIHYPRSTPQMPDIIQAKD 65
Db 12 MNLVYVFLITLISCVGNWYDRAIKINDORILLSGSIHYPRSTPQMPDIIQAKD 71
Qy 66 GGLDVITYVFWNGHEPSGKYNFEGRYDLVRFIMVQAGLYNLRIGPYVCAEWNFGG 125
Db 72 SOLDVITYVFWNGHEPSGKYNFEGRYDLVRFIMVQAGLYNLRIGPYVCAEWNFGG 131
Qy 126 FVWLKYYVPGMEFTNNOPFKVAGQFVOKIVNMKSENLFESQGGPIIMAGIENEYGPV 185
Db 132 FVWLKYYVPGMEFTNDGPFKFKQVFTTKIVDMKAELFHWGGPILILAGIENEYGPV 191
Qy 186 EWEIGAPGKATYTKAAQMAVGLKGVPMCKQOE-DAPDPVIDTCNGFYCEGFENKPKYK 244
Db 192 EWEIGAPGKATYTKAAQMAVGLKGVPMCKQOE-DAPDPVIDTCNGFYCEGFENKPKYK 251
Qy 245 PKMTWETWYTGKGGIPQPAEDIAFVSARFVQNNGSFFNYMYHGGNFGRFTSSGL 304
Db 252 PKMTWETWYTGKGGIPQPAEDIAFVSARFVQNNGSFFNYMYHGGNFGRFTSSGL 310
Qy 305 FIATSYDAPLDEYGLNKEPKYGHLDLHKAIKLSEPALVSSYAATVSLGSQNAHYVR 364
Db 311 FVSTSYDAPLDEYGLNKEPKYGHLDLHKAIKLSEPALVSSYAATVSLGSQNAHYVR 370
Qy 365 SKGACAAFLSNYSRYSVKYTFQNRPNLPPWSTSLPDCCKTAVYNTAQVNSOSSST-- 422
Db 371 SNKGYKPKMTWETWYTGKGGIPQPAEDIAFVSARFVQNNGSFFNYMYHGGNF-- 297

Qy 423 KMTAGGGLSMQSYNEETPTADSDTLTANGLEQKNVTRDSSDYLTWYMTNVIASNEGF 482
Db 431 KMTPTVSNLWQSYDEVPADSPGTFREKLYEQINMTWDSYLTWYMTDVLVDGNEGF 490
Qy 483 LKNGKDPVLTVNSAGHVLHVFNGLSGTVYCTLDNPKLTYSGNKLKAGINKISLLSVS 542
Db 491 LKKGDEPWLTVNSAGHVLHVFNGLSGTVYCTLDNPKLTYSGNKLKAGINKISLLSVS 550
Qy 543 VGLPNVGVHYDTWNAAGVLPVTLGSLNEGSRNLAKQKWSYKVLKGESLSLSLSSGSSV 602
Db 551 VGLANVGVHFFERYNGVLPVTLGSLNEGSRNLAKQKWSYKVLKGESLSLSLSSGSSV 610
Qy 603 EWRGSLMAQKQPLTWYKATFNAPGGNDPLALDMSMGKQIWIINGEGYGRHWPGYIAQ 662
Db 611 QW---GPPAWKQPLVWYKATFNAPGGNDPLALDMSMGKQIWIINGEGYGRHWPGYIAQ 667
Qy 663 DCS-KCSYAGTFNEKKCQTNCCQPSQRYVHPRSWLKPSCNLLVVFEEWGGNPTGISLVR 721
Db 668 SCNDNCNAGTYTETKCLSDCGKSKQKWHVPRSWLQPRGNLLVVFEEWGGNPTGISLVR 727
Qy 722 RS 723
Db 728 RT 729
RESULT 4
US-09-436-605-2
; Sequence 2, Application US/09436605
; Patent No. 6140115
; GENERAL INFORMATION:
; APPLICANT: Kolodny, Edwin, H.
; APPLICANT: Wang, Zhao-Hui
; APPLICANT: Raghavan, Srinivasa,
; APPLICANT: Seng, Baijin
; TITLE OF INVENTION: The Canine (-Galactosidase Gene and Gml-Gangliosidosis
; FILE REFERENCE: D6273
; CURRENT APPLICATION NUMBER: US/09/436,605
; CURRENT FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 25
; SEQ ID NO 2
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Canis familiaris
; FEATURE:
; NAME/KEY: MUTATION
; LOCATION: 60
; OTHER INFORMATION: Arg to His mutation in R60H allele.
; OTHER INFORMATION: Portuguese Water dog (-Galactosidase Protein
US-09-436-605-2

Query Match 12.4%; Score 485; DB 4; Length 668;
Best Local Similarity 26.0%; Pred. No. 7.1e-35;
Matches 179; Conservative 83; Mismatches 244; Indels 182; Gaps 24;
Qy 24 SVSYDDRAIIINGKRKTLISGSIHYPRSTPQMPDIIQAKDGLDVTETVYVFWNGHEPS 83
Db 34 TIDYSHNFKDQGFPRVYISGSIHYPRSTPQMPDIIQAKDGLDVTETVYVFWNGHEPS 93
Qy 84 PKYVNEGYDILVRFKVQVQAGLYNLRIGPYVCAEWNFGGPIIMAGIENEYGPV 143
Db 94 PQYQFSGEQDVEYFPIKLAHELGLLVLRPGPYICAEDMDGMLPAWLLKELIILRSSDP 153
Qy 144 PKVAMQGFQKIVNMKSENLFESQGGPIIMAGIENEYGPV---PVWEI----- 189
Db 154 DYLAADVADKWLGVLLPKMKP--LLYQNGGPIITMVENEYGYFTCDYDLRFLKQLFHHH 211
Qy 190 -----GAPGKATYTKAAQMAVGLKGVPMCKQOEADPDPVIDTCNGFYCEGF 238
Db 212 LGNDVLLFTDGCANEKFLQCGALQ---GLYATVDF-----GPGANITAAFIQ--R 257
Qy 239 PNKPYKPKMTWETWYTGKGGIPQPAEDIAFVSARFVQNNGSFFNYMYHGGNF-- 297

Db 258 KSEPKGPLYNSEFTYGLDHWGPHSTVTEVVASSL-HDIIAHGANVNLWYFVIGTNPFA 316
QY 298 ----GRTSSGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKA-----IKLSEPALVS 346
Db 317 YWNGANMPYQAQPTSYDYDAPLSEAGDLTE-KYFALREVIRKEKVPCEGFIPSTPKFAY 375
QY 347 SYAATVSLGSNOQDAHYRYSKSGACAAFLSNYDSRYSVKVTQNRPNYLPWISILPDCK 406
Db 376 GRVALKLLKTBEEALNVLCPGPG-----INSLYPLTEIQVKQVFGFVWYRTTLPDQCS 428
QY 407 TAVYNTAQNVSOSSTKTKPAGGGLSWQSYNEETPTADDSDLTANGLWEQKNVTRDSD 466
Db 429 DP-----TPLSSPLS-----GVHDRAVSVVG-- 450
QY 467 YLWYMTNVNIASNEGFLLKNGKDPYLTVMAGHVLHVFNKGLSGTVYGLDNPCKLTYSGN 526
Db 451 -----VPOG-----VMERSNVITLITGKAGAT----- 473
QY 527 VKLRAGINKISLSVSVGLPNVGVHYDTWNAGVLPVTLVSLGSLNEGSRLNAKOKW----- 580
Db 474 -----LDLLVENMGRVNYGRYINDFK-GLISNLTL-----GSSILT--NWMIFPLN 516
QY 581 -----SYKVLKLGESLSLH--SLGSSSVVWVRSGLMAQKQPLTWYKATFNAPGNDPL 632
Db 517 TEDAVRSHLGGWHGPNNGRHRDXTFAHRSS-----NYTLPA-----FYMGNFSIPSGIDPL 566
QY 633 ALD-----MASMGKGQIWIINGEGVGRHWP 656
Db 567 PDDTFIQPPGWTQGWINGFNGLGRYP 594

RESULT 5

US-08-596-985-2
; Sequence 2, Application US/08596985
; Patent No. 5736374
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Hucul, John A.
; APPLICANT: Ward, Michael
; TITLE OF INVENTION: Increased Production of
; TITLE OF INVENTION: Beta-galactosidase in Aspergillus oryzae
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: South San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/596.985
; FILING DATE: 05-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/267,631
; FILING DATE: 29-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC250
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-596-985-2
Query Match 10.1%; Score 396.5; DB 1; Length 911;
Best Local Similarity 23.0%; Pred. No. 1.1e-26;
Matches 197; Conservative 98; Mismatches 316; Indels 245; Gaps 36;
QY 25 VSYDDRAIITNGKRLTISGSIH-VPRSPQMPDLOKAKDGGDLVETIYVFWNGHEPS 83
Db 46 VTWDDKSLFINGERIMLFSEGVHPFLPVLPSLWLDIFKIRALGFNCVSYFIDWALLEGK 105
QY 84 PGKYFEGEDYLVREIKVMQVAGLYNLRIGPYVCAEMNFGGFPVWMLKYVPGNEFTNQ 143
Db 106 PGDYRAEGIFALEPPEFDAKAAGIVLIARPGSYINAESVGGGPGWLQVRNG-TLRSDE 164
QY 144 PKFVAMQGVQKIVNMKSENLFESQGGPIIMAOITENEGYGVWEIGAPKAYT-----K 198
Db 165 PFLKATDNYIANAAAVAKAQI--TNGGPIVLYQENESYSG-----GCCGVKYTADYMQ 217
QY 199 WAAQMAVGLKTGPVIMCKQEDAPD-----PVIDTCN-----GFYC-----EG 236
Db 218 YVMDQARKADIIVPFI--SNDASPSGHNAPGSGTGAVDIYGHDSPLGPDCAINPWPBG 275
QY 237 FRPN-----KPYKPKMTTEVWTGWTKFGGPIQORPAEDIAFVSAR-EVQNGNSP- 285
Db 276 KLDPNFTLHLRQSPSAPYSLLEFOAGAFDPWGGPGFEKCYALVNHFSRVFYRNDLSFG 335
QY 286 ---FNYMYHGTNFGRTS-SGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSE 341
Db 336 VSTFNLYMTFGTNGNLGHPGGY--TSYDYGSPITETRNVTREKYSIDIKLANEFKASP 393
QY 342 PALV-----SSYAAVTSI-----GSNOEAHVYRSKSGACAA- 372
Db 394 SYLTATPNRLTGVYDTSDLAVTPLIGDSPGSGFFVVRHTDYSSQESSTYKLUKPTSGN 453
QY 373 -----FLSNYDSRYSVKVTQNRPNYLPWISILPCKTAVYNTAQNVSQSSI 422
Db 454 LTIPQLEGTLSLNGRDSKTHVV-----DYNVSGTNI-----IYSTAEV----- 491
QY 423 KMTAGGGLSWQSY-----NEEPTADDSLTTLTANGLWEQKNVTRDSDYL 468
Db 492 -----FTWKKFDGNKVLVLYGGPKKEHHELAIAASKSNVTIIEGSDSIVSTRKSSVI 543
QY 469 --WYMTNVNIASNEGFLLKNGKDPYLTVMAGHVLHVFNKGLSGTVYGLDNPCKLTYSGN 526
Db 544 IGW-----DVSSTRIVQVGD-----LRVFLLGKNSAYNYWVPELP--TEGTS 584
QY 527 VKLRAGINKISLSVSVGLPNVGVHYD-----TNAGVLPVTLVSLGSLNEGSRLN----- 575
Db 585 PGFSTSKTTASSIIVKAGYLLRGALHDGADLHLTADFNATTPTEVIGAPTGAKNLNVNGE 644
QY 576 -----AKQWSYKVLKGESLSLHSL-----SCSSSVWVWVRSGLMAQK 613
Db 645 KASHTVDXNGINWSEVKAAPKILPGLKDLDMKLDLTLPETKSYDDSAWYSADLPKTK 704
QY 614 ---QPLTW-----YKATFNAPGNDPLAL--DMASMGKGQIWIINGEGV 651
Db 705 NTHRPLDTPTSLSYSDYGFHGYLIRGHFVANGKESEFLINTQGSAGFSSVWLNLEYL 764
QY 652 GRHWPYIAQGDCKSKCYAGTFNEKKCQTNCGQPSQRWYH-----PRSWLKPSG 701
Db 765 G-SWTG-----ADYTMDSNSTYKLSQLESN-----YHVTIVIDLNLGLENW----- 806
QY 702 NLLVVFEEWGNPTGI 717
Db 807 ---TVGEETMKNPRGI 819
RESULT 6
US-08-551-459-4
; Sequence 4, Application US/08551459

Patent No. 5821350
; GENERAL INFORMATION:
; APPLICANT: Huang, Yue
; APPLICANT: Karatzas, Costas N.
; APPLICANT: Lazaris-Karatzas, Anthoula
; APPLICANT: Delaquis, Annick
; TITLE OF INVENTION: ASPERGILLUS NIGER BETA GALACTOSIDASE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/551,459
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06632/005001
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1007 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-551-459-4

Query Match 10.1%; Score 395; DB 2; Length 1007;
Best Local Similarity 23.4%; Pred. No. 1.8e-26;
Matches 187; Conservative 103; Mismatches 308; Indels 200; Gaps 32;

Qy 25 VSYDDRAIIINGKRIKILSGSIHPRSTP--QMPPDLLOKAKDGLDVIETVFNHGHEP 82
Db 46 VTWDDKSLFINGERIMIFSGEFH-PFRLPVKELQDLDFQKVKALGFNCVSYVDWALVEG 104
Qy 83 SPGKYNFEGRYDLVRFKMWQRAGLYVNLRIQPYVCAEWNFGGFPVWLKYVPGMFEFTNN 142
Db 105 KPGEYRAGDIEFDLPPFDDAASEAGIYLARPGYINAESSGGGFPGLQRYNG-TLRSSD 163
Qy 143 QPFVKVMOGQFVQKIVNMKMSLENFESQGGPIIMAEIENYGPVWEIGAPGKAYTKWAAQ 202
Db 164 KAYLDATDNVYSHVAATAIKYQI--TNGGPITLYOPENEYTSGCCGVFPDPVYMQYVED 221
Qy 203 MAVGLKTGVPI---MCKQEDAPP---VIDTCN-----GFYC-----BGF 237
Db 222 QARNAGVVIPLINNDASAGSNAPATGKGAVDIYGHDSYPLGFCANFTVMPSGDLPTNF 281
Qy 238 R----PNKPYKPKMWTEVMTGWYTKFGGPIQRPADIAFSVAR-FVONNGSF----FNY 288
Db 282 RTLHLEQSPPTPYAIVEFGGSDYDPPGPGFACSELNNFEFVYKNDIFSQTAINWL 341
Qy 289 YMYHGGTNFGRTS---SGLFIATSYDYDAPLDEYGLLNEPKYCHLRDLHLKAIKLSEPALVS 346
Db 342 YMIFGTTWNGNLGYPNGY---TSYDYGSAVYTESRNTREKYSELKLLGNFAKVPSPGYLTA 398
Qy 347 S-----YAAVTSLG-----SQEAAHVYRSKSGACAA----- 372
Db 399 SPGNLTTSGYADTTDLTVTLPLGNSTGSGFFVVRHSDYSSESTSYKLRPLTSAGSVTIPO 458

Qy 373 -----FLSNYDSRYSVKVTQNRPNLPPWSISILPDCKTAVYNTAQV----- 415
Db 459 LCGTITLNGRDSKIHV-----TDYVSGTNI-----YSTAEVFTWKKFADGKVL 503
Qy 416 -----NSQSSSIKMTPAGGGLSQWSYNEETPTADSDT-----LTFANGLWEQK 458
Db 504 VLYGGAGEHHELAISTKSNVTIEGSESGISKQTSSSVVWGVWDVSTTRRIIOVGDLKIL 563
Qy 459 NVTROSSDYLYWYMTNVNNTASNEGFLKNGKDPYLYTMSAGHVLFVFNKLSGTVYGTLDN 518
Db 564 LLDRNSAYNWVWPOLATDGTSPGFSTPEKVASSIIVKAGYLVR-----TAY----- 609
Qy 519 PKLTYSGNVKLRAGINKISLLSVSGLPNVGVHYDTWAGVLGPTVTLNGLNGSRNLAKQ 578
Db 610 --LKSGG-LYLTADFNAATSVVEV-IGVPSTAK-----NLFINGDKTSHVDKNG----- 654
Qy 579 KWSYKVGKLGESLSLSLSG-----SSSVE---WVRGSLMAQKQPL----- 616
Db 655 IWSATVDYNAPDISLSPLKDLWKYVDTLPEITQSSYDDSLWPAADLKQTKNLTLSLTPT 714
Qy 617 -----TWYKATFNAPGGNDPLALDM--ASMCKGOIWIENGEGVGRHWPYIAQ 661
Db 715 SLYSSDYGFHTGYLLYRGHFTATGNESTFAIDTQGSAPGSSVWLNGLTYLG-SWTGLYAN 773
Qy 662 GDCSKCSYAGTFNEKKCQ 679
Db 774 SD-----YNATYNLPQLQ 786

RESULT 7
US-08-840-062-5
; Sequence 5, Application US/08840062
; Patent No. 6117977
; GENERAL INFORMATION:
; APPLICANT: LASKY, LAURENCE A.
; TITLE OF INVENTION: TYPE C LECTINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,062
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: PL019R1
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1455 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-840-062-5

Query Match 3.7%; Score 146.5; DB 3; Length 1455;
Best Local Similarity 20.1%; Pred. No. 0.00076;

Matches	139;	Conservative	73;	Mismatches	229;	Indels	249;	Gaps	36;
Qy	121	WNFGGFPWL-----RYVPGMEFRTNQPPKVA-----QGFQKIVNMKSKENLFESQ	169						
Db	507	WKRHGFCYLLIGSLSTFTDANHCTNEKAYLTIVEDRYEQAFSLVLG-LRPEKYFWTG	565						
Qy	170	GGPIIQAQIENEXGPVWEIGAPGAKATYKAAQM-----AVGLKTGVP---W-IMKQE	219						
Db	566	-----LSDVQNK-GTFRTWVDEQVQ-QTFHWADMPGRKAGCVAMKTGVAGLWDLVLSCEE	618						
Qy	220	DAPDVIDTCNGFCEGRPNPKPKPKMTWETWGTGKFGGPIQPQPAEDIAESVARFV	279						
Db	619	KAK-----FVC-----KHWAEGVTRPPEPTTTPPKPC-----	646						
Qy	280	QNNGSFFNYMYHGGTINFGRTSSGLFIATSYDYDAPLDEYGLLNPYKYLHDLHKAIKL	339						
Db	647	-----ENWGTTSK-----TSMCF--KLYAKGKHKKWTFESRDCFKALG-	683						
Qy	340	SEPALVSS-----YAAVTSILGSNQEAH-----VYRSKSGACAAFLNYSYSYKVTFFQ	388						
Db	684	GELASIKSDEQQVIWRLITSSSGSYHELFWGLTYGSPS-----EGFTWSDGSPVSYE	736						
Qy	389	NRPNLNPWSSISILPDCKTAVYNTAQNVSQSSSTKMTPGAGGLSQSYN-----	437						
Db	737	NWAYGEP-----NNYQNVEYCGELKGPD--GMSWNDINCEHLNNWICQI	778						
Qy	438	-----EETPTADDSDTLTANG-----LWEQKNVTRDSS-----	465						
Db	779	QKGYTLPEPTPAPQDPPVPTADGWVIYKQYYSKEKETMDNARFCNKNFGDLATIK	838						
Qy	466	-----DYLWNTMNTNIAISNEGFLNGKDPY-LTYMSAGHVLHFVFNPKLSVTYGYTLDPN	519						
Db	839	SESEKFLWKYINKN-----GGOSPFIYGLMSIDKKFIWMDGSKYDFAVATGEP	889						
Qy	520	KLTYSGNVKLRAGLNKISLSSVGLPNVGV---HYDTWNAVL--GPVTLUSGLNE----	570						
Db	890	NFANDDENCVTWYNTSGFWNDINCGYNNFICQRHNSINATAMPTPTTTPGCKEGWHL	949						
Qy	571	-----GSRNLAKOKWSYK-----VGLGESLSLSLSSSSSVWVRGSLMAQKOPLT	617						
Db	950	YKNCKFIKIFGANEKSKWDQARQCKGLKNLVSIEN-----AQEQAFV	994						
Qy	618	WY---KATFNA-PGQNDPLALDMSMGKQITWINGEV-----GRHWPGYIAQDCSKCS	668						
Db	995	TYHMRDSTFNWATGLNDINAEHMF-----LWTAGQGVHYTNWCKGYPG-----GRRSSL	1044						
Qy	669	Y-----AGTFNEKKCQNCG	683						
Db	1045	YEDADCYVVYIGGNSREAGTWMDDTCDCKSQG	1074						

```

RESULT      8
US-08-687-372-1
; Sequence 1, Application US/08687372
; Patent No. 5859344
;
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R
; APPLICANT: HOLT, KAREN A
; APPLICANT: PICARD, SYLVIE
; APPLICANT: SCHUCH, WOLFGANG W
; APPLICANT: CAREY, ANNETTE T
; APPLICANT: SEYMOUR, GRAHAM B
; APPLICANT: TUCKER, GREGORY A
;
; TITLE OF INVENTION: MODIFIED FRO
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUS
; STREET: 1100 New York Avenue,
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
;
; COMPUTER READABLE FORM:

```

```

? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/687,372
? FILING DATE: 2-AUG-1996
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: GB 9320930.2
? FILING DATE: 12-OCT-1993
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: PCT/GB94/02203
? FILING DATE: 10-OCT-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: KOKULIS, PAUL N.
? REGISTRATION NUMBER: 16,773
? REFERENCE/DOCKET NUMBER: 224239/SEE37803X/UST
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (202) 861-3000
? TELEFAX: (202) 822-0944
? TELEX: 6714627 CUSH
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 30 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
;
US-08-687-372-1

Query Match          3 6%; Score 141; DB 2; Length 30;
Best Local Similarity 93.3%; Pred. No. 3.8e-06;
Matches 28; Conservative 0; Mismatches 2; Indels

QY 24 SVSYDDRAIIIGKRKILISGSIHYPRSTP 53
Db 1 SVSYDDRAIIIGKRKILISGSIHYPRKYP 30
      |||||
      |

RESULT 9
US-08-687-372-7
; Sequence 7, Application US/08687372
; Patent No. 5859344
; GENERAL INFORMATION:
; APPLICANT: BIRD, COLIN R
; APPLICANT: HOLT, KAREN A
; APPLICANT: PICARD, SYLVIE
; APPLICANT: SCHUCH, WOLFGANG W
; APPLICANT: CAREY, ANNETTE T
; APPLICANT: SEYMOUR, GRAHAM B
; APPLICANT: TUCKER, GREGORY A.
; TITLE OF INVENTION: MODIFIED FRUIT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,372
; FILING DATE: 2-AUG-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9320930.2

```


;; FILING DATE: 12-OCT-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/GB94/02203
;; FILING DATE: 10-OCT-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: KOKULIS, PAUL N.
;; REGISTRATION NUMBER: 16,773
;; REFERENCE/DOCKET NUMBER: 224239/SEE37803X/UST
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 861-3000
;; TELEFAX: (202) 822-0944
;; TELEX: 6714627 CUSH
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 30 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-687-372-7

Query Match 3.5%; Score 137; DB 2; Length 30;
Best Local Similarity 86.7%; Pred. No. 8.7e-06;
Matches 26; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 24 SVSYDRAIIINGKRKILISGSIHYPRSTP 53
DB 1 SVSYDHKALIIVGQRKILISGSIHYPRSTP 30

RESULT 10
US-08-467-023-2
; Sequence 2, Application US/08467023
; Patent No. 6090386
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.;
; APPLICANT: Pollock, Joanne;
; APPLICANT: Bond, Julian F.;
; APPLICANT: Garman, Richard D;
; APPLICANT: Kuo, Mei-Chang;
; APPLICANT: Yeung, Siu-mei H.;
; APPLICANT: Brauer, Andrew;
; APPLICANT: Exley, Mark A.;
; APPLICANT: Powers, Steven P.
; TITLE OF INVENTION: Allergenic Proteins And Peptides From
; TITLE OF INVENTION: Japanese Cedar Pollen
; NUMBER OF SEQUENCES: 261
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.
; STREET: 610 Lincoln St
; CITY: Waltham
; STATE: MA
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467.023
; FILING DATE: June 6, 1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/350,225
; FILING DATE: December 6, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Jane E. Remillard
; REGISTRATION NUMBER: 38,872
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400

;; TELEFAX: (617) 227-5941
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 374 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-467-023-2

Query Match 3.2%; Score 125; DB 3; Length 374;
Best Local Similarity 21.7%; Pred. No. 0.0069;
Matches 89; Conservative 38; Mismatches 139; Indels 144; Gaps 21;

QY 276 AREVQNGSFNMYVHGTFNFGRTSSGLFIATSYDYDAPLDEYGLNPKYGHRLDL-- 333
DB 33 SNWAQRNKLADCAVGFSGSSTMGKGGLYTVTNSDDDPVNPAPGFL---RYGATRDRL 89
QY 334 -----HKAIKLSEPALVSSYAAVTSLSGNOEAAHYVRSKSGACAAFLSNYSRYSVKVTF 387
DB 90 WIIFSGNMNIKLMKPMYIAGYKTFDGRG---AQYVIGNGGPC-----VF 130
QY 388 QNRPNLPPWSISILPDCKTAVYNTAQVNSQSSSKMTMPAGGLSKWSQSYNEETPTADSD 447
DB 131 IKRVSNIHGL-YLYGCSSTVLGNVLN-----ESFGVEPVHPQDGD 172
QY 448 TL---TANGLWEQKNVTRDSDYLWYMTNNTA---SNEGFLKNGKDPYLTVMNSAGHVHL 501
DB 173 ALTLRTATNINIDHNSFSNSDGLVDVTLTSTGVTSNNLFFNHHK-----VLLGH--- 224
QY 502 VEVNGKLSGTVYGTLDNPKLTYSG--NVKLRAGINKISLLSVSLPNVG----- 549
DB 225 -----DD---AYSDDKSMKVTVAFNQFG-----PNCQRMPRARYGL 258
QY 550 VH-----YDTWAGVLG-----PVTLSGLNEGSRNLAKQKWSYKVLKGBESLSLSLSSGS 600
DB 259 VHVANNYDPWTIYAIGGSSNPTILS---EGNSFTAPNE-SYK---KQVTIRIGCKTSSS 311
QY 601 SVEWV-----RGLSMAOKOPLTWYKATFNAPGN 629
DB 312 CSNWWQSTQDVFYNGAYFVSSGKYEGGNIYTKKE-----AFNVENG 354

RESULT 11
US-07-977-451-4
; Sequence 4, Application US/07977451
; Patent No. 5270458
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varlick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977.451
; FILING DATE: 19921119
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US UNASSIGNED
; FILING DATE: 12-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397

```

; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELEPHONE: 212-645-2054
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; LENGTH: 993 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-977-451-4

Query Match 3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No. 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;

QY 61 QKAKGGGLDVITYVFNWGHSPGKYNFEGYDLVRFKMKVQKAGLYNLRIGPVVCAE 120
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 394 QKGLDNGYSISK---FCN-HKHQPGGEYIFHAENDDAQFTKM-----FTLNIRKPKQVLAE 444
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 121 WNFQ-----GPPVWLKYPVGMFEFTNQPKVAMQGFQKIVNMKSENLFESOGPI 173
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 ASASQSCFSDGYP-----LPSWTWKKCD-----KSPNCTE----- 476
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 174 IMAQIENEGPVWEITGAPGKAYTKWAQMAVGLKTGPWIMCKQEDAPDPVIDTCN--- 230
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 477 -----EITEGWNRKANRKGQVGSSTLNMSEAIGKFLVK-----CCAYNS 519
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 -GFYCGFRPNPKPKMWTETWGTWYTKFGGPIQORPAEDIAFSVARVQNGSFF--- 286
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 520 LCTSCETILLNSP-----GPPF-----FIQDNISFYATI 548
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 -----NYMYHGGTNE-----GRTSSGLFIA--TSYDYDAPLD--- 317
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 GVCLLFIVVTLILCHIKYKKQRYESQLOMQVQVTGSSDNEYFYVDFREYDLKWEFPRE 608
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 318 --EYG-----LLNEPKYGHLR---DLHKAIKL-----SEPALVSSYAAVTS 354
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 609 NLEFGKVLGSGAFGKVMNATAGISKTVSIOVAVKMLKEKADSSREALMSELKMTOL 668
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 GSNQ-----EAHYRKSAGCAAFLSNY-----DSRYSVKVT 386
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 669 GSHENIVNLLGACTLSGPIYLIFEYCCYDGLLNLRKREKFRHTWTETFEKHNFSEYPT 728
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 387 FQNRPNLPWISILPDCKTAVYNTAQVNSQSSSKTKTPAGGGLSQWSYNEETPADDS 446
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 729 FQSHPNSSMPGSGREV-----QIHPDSQDI-----SGLHGSFHSFSE---DEI 766
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

447 DTLTANGLWEQNV-TRDSSDYLWYMTNNTNVIASNEGFLKNGKDPYLTVMAGHVLHVFVN 505
: : : : : : : : : : : : : : : : : : : : : : : : : :
767 EVENQKRLEEDNLVLTFTEDLLCFA--YQVAKGMEFLE-FKSCVHRDLAARNVL--VTH 821
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 506 GKL-----SGTVYGTLDNPKLTYSGNVKLRAGINKISLISVSVGLPNVGVHYDTWAGVL 560
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 GKVKICDFGLARDIMSDSNYVVRGNARL--PVKWMAPESLFEGI--YTIKSDVWSYGIL 877
||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 12
US-08-252-517-4
; Sequence 4, Application US/08252517
; Patent No. 5548065
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ihor R.
; TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL
; TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,517
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,451
; FILING DATE: 19-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/906,397
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/05401
; FILING DATE: 26-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: TW 81102961
; FILING DATE: 15-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US92/02750
; FILING DATE: 02-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/813,593
; FILING DATE: 24-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,065
; FILING DATE: 15-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/728,913
; FILING DATE: 28-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/679,666
; FILING DATE: 02-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TYPE: amino acid
```



```
; NAME: Feit, Irving N.
; REGISTRATION NUMBER: 28,601
; REFERENCE/DOCKET NUMBER: LEM-3-7P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-645-1405
; TELEFAX: 212-645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-601-891-4

Query Match      3.0%; Score 118; DB 1; Length 993;
Best Local Similarity 18.5%; Pred. No 0.15;
Matches 111; Conservative 74; Mismatches 199; Indels 216; Gaps 30;

QY 61 QKAKDGLDVIETVFWNGHEPSGKYNFEGRYDLVRFIMQVORAGLYVNLRIQPVYCAE 120
Db   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
394 QKGLDNGYSISK---FCN-HKHQGEYIFHAENDDAQFTKM-----FTLNIRKQVLA 444
QY 121 WNFQ-----GPPVWLKYVPGHEFTNNOPFKVAMQGFVKIVNMKSENLFESOGGPI 173
Db   : : : : : : : : : : : : : : : : : : : : : : : : : :
445 ASASQASCSFSDGYP-----LPSWTWKKCD-----KSPNCTE----- 476
QY 174 IMAIENEGYGPVEIGTGAQKAYTKWAAQMAVGLKTGVPWIMCKQEDAPDPVIDTCN--- 230
Db   | | | | | : : : : : : : : : : : : : : : : : : : : :
477 -----ETEGVWNRKANRKFQGVSSSTLNMSAIGFLVK-----CCAYNS 519
QY 231 -GFYCEGFRPNKPKPKMTEVWTGTYTKFGGPIQORPAEDIAFVSARFVQNNGSFF--- 286
Db   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
520 LGTSCETILLNSP-----GFPF-----FIQDNISFYATI 548
QY 287 -----NYMYHGGTNP-----GRTSSGLFIA--TSYDYDAPLD----- 317
Db   : : : : : : : : : : : : : : : : : : : : : : : : : :
549 GVCLEFIVVLTLLICHKKYKQFRIESQLQMVQVTGSSDNEYFYVDFREYEDLKWEFPRE 608
QY 318 --EYG-----LLNEPKYGHLR---DLHKAIKL-----SEPALVSSYAAVTSL 354
Db   ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
609 NLEFGKVLGSGAPGKYVNNATAYGISTGYSIQVAVAKLKEKADSSEREALMSLKMMTQL 668
QY 355 GSNQ~-----BAHYRSKSGACAFLSNY-----DSRYSVKVT 386
Db   ||: : : ||: : : ||: : : ||: : : ||: : : ||: : : ||:
669 GSHENIVNLLGACTLSGPIYLIFEYCCYDGLLNLRSKREKPHRTWTEIFKEHNEFYPT 728
QY 387 FQNPYNLPWISILPDCKTAVYNTAQVNSQSSSTKMTFAGGGLSWQSYNEETPTADDS 446
Db   ||: : : ||: : : ||: : : ||: : : ||: : : ||: : : ||:
729 FOSHPNSSMPGSRV-----QIHPDSQI-----SGLHNSFHSE----DEI 766
QY 447 DTLTANGLWEQKNV-TRDSSDYLWYNTNVIASNEGFLKNGKDPYLTVMASAGHVLHVFVN 505
Db   : | : : : : | : : : : | : : : : | : : : : | : : : : |
767 EYENQRLREEDDLNVLTEDDLCAFA--YQVAKGMFELE--FKSCVHRDLAARNVL--VTH 821
QY 506 GKIL-----SGTVYGTLDNPKLAYSGNVKLRAGINKISLLSVSGVGLPVGVDYDTWAGVL 560
Db   ||: : : ||: : : ||: : : ||: : : ||: : : ||: : : ||:
822 GKVKICDFGLARDINSDSNYYVVRGNARL--PVKWWAPESLFEGL--YTIKSDVWSYGIL 877
```

Search completed: June 24, 2002, 20:37:36
Job time: 222 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2002, 20:35:45 ; Search time 35.44 Seconds
(without alignments)
2269.115 Million cell updates/sec

Title: US-09-701-868-11
Perfect score: 3924
Sequence: 1 MLRTNVLVLLVLCILDFSS.....VPEWGGNPTGISLVRRSR 724

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : **US-09-701-868-11**
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3924	100.0	724	21	AA1980
2	2941	74.9	838	21	AA1981
3	2922.5	74.5	835	21	AA1982
4	2920	74.4	838	16	AA1983
5	2896	73.8	730	16	AA1984
6	2808.5	71.6	724	21	AA1985
7	2645.5	67.4	666	21	AA1986
8	2608	66.5	853	21	AA1987
9	2608	66.5	856	21	AA1988
10	2572	65.5	847	21	AA1989
11	2572	65.5	850	21	AA1990

12	2572	65.5	850	21	AAG39970	Arabidopsis thalia
13	2572	65.5	853	21	AAG31160	Arabidopsis thalia
14	2482.5	63.3	794	21	AAG49424	Arabidopsis thalia
15	2446.5	62.3	788	21	AAG39972	Arabidopsis thalia
16	2446.5	62.3	791	21	AAG31162	Arabidopsis thalia
17	2240.5	57.1	585	21	AAG49311	Arabidopsis thalia
18	2183.5	55.6	857	17	AAR95287	Nasturtium xyloglu
19	2077	52.9	870	21	AA144309	Tomato beta galact
20	1918	48.9	887	21	AA144304	Tomato beta galact
21	978	24.9	251	21	AA144307	Tomato beta galact
22	970	24.7	279	21	AAG23589	Arabidopsis thalia
23	951	24.2	274	21	AAG23590	Arabidopsis thalia
24	927	23.6	249	21	AA144308	Tomato beta galact
25	738	18.8	295	21	AAG27645	Arabidopsis thalia
26	622	15.9	252	21	AAG19689	Arabidopsis thalia
27	536	13.7	636	22	AA193378	Human polypeptide,
28	535	13.6	636	21	AA12157	Hydrophobic domain
29	531	13.5	636	20	AA113374	Amino acid sequenc
30	531	13.5	636	22	AA129034	Human PRO polypept
31	531	13.5	636	22	AA180242	Human PRO236 prote
32	527	13.4	636	22	AA193429	Human polypeptide,
33	509.5	13.0	637	21	AA171116	Human Hydrolyase pr
34	496	12.6	696	21	AA156675	Human prostate can
35	495	12.6	677	19	AA170900	Human beta galacto
36	486	12.4	215	21	AAG27646	Arabidopsis thalia
37	485	12.4	668	22	AA136888	Portuguese Water d
38	484.5	12.3	552	22	AA193931	Human polypeptide,
39	483.5	12.3	654	22	AA129036	Human PRO polypept
40	483.5	12.3	654	22	AA106593	Human protein havi
41	483.5	12.3	654	22	AA180243	Human PRO262 prote
42	473	12.1	598	20	AA124312	Lactobacillus case
43	470	12.0	123	21	AAG35971	Arabidopsis thalia
44	467.5	11.9	598	20	AA124313	Lactobacillus case
45	461	11.7	687	22	AA164229	Drosophila melanog

ALIGNMENTS

RESULT 1
AA144306
ID AA144306 standard; protein; 724 AA.
AC AA144306;
XX
DT 29-FEB-2000 (first entry)
XX
DE Tomato beta galactosidase-4.
XX
KW Tomato beta galactosidase-4; TBG; Rutgers tomato plant; pectin;
KW fruit softening; beta galactosidase II protein; biofilm;
KW transgenic plant; protoplast isolation.
XX
OS Lycopersicon esculentum.
XX
FH Key Location/Qualifiers
FT Modified-site 282
FT Modified-site /note= "Possible N- glycosylation site"
FT Modified-site 459
FT Modified-site /note= "Possible N-glycosylation site"
FT Modified-site 713
FT Cleavage-site /note= "Possible N- glycosylation site"
FT Cleavage-site 23..24
FT /label= Leader_sequence_cleavage_site
XX
XX WO9964564-Al.
XX
XX 16-DEC-1999.
XX
XX 08-JUN-1999; 99WO-US12697.
XX
XX 09-JUN-1998; 98US-0088805.
XX

PA (USDA) US DEPT OF AGRICULTURE.

AA
PT
Gross KC
Smith Df.

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

DR N-PSDB; AAZ29341.

PT New beta-galactosidases, used to prepare transgenic plants with altered
PT fruit ripening -

PS Claim 1; Fig 2; 85pp; English.

The present sequence is tomato beta galactosidase-4 (TBG-4) encoded by a cDNA derived from breaker, turning and pink fruit pericarp from 'Rutgers' tomato plants. This hydrolyses terminal non-reducing beta-D-galactosyl residues from beta-D-galactosides leading to loss of tissue integrity and fruit softening. This is used for modifying cell wall metabolism and controlling ripening of fruit by altering activity of beta galactosidase in protein. Pectin with reduced galactosyl content is produced for use in biofilms or solutions. Transgenic plants with altered fruit ripening are produced by introducing DNA constructs comprising TBG cDNA. TBG forms a component of an enzyme mixture used to isolate protoplasts.

Sequence 724 AA:

```
Query Match      100.0%; Score 3924; DB 21; Length 724;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 724: Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

1 MLRTNVLLLVICLLDEFFSSVKASVSYDDRAIIINGKRKILISGSIHYPRSTPOMWPDLI 60

pb j m]rtov]lllv]c]l]dfffssvkaevsvd]d]ra]i]nq]rk]l]i]ssg]i]b]u]n]r]g]t]n]a]m]w]d]l]f]60

0v 61 0KAKDGGI:DVTFYVFWNGHEPSPCKYNFEGRYDI.VBETKMWOBAGI.VVNI.RIGPYVCAE 120

61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 10

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

100

[illegible][illegible]

241 KPYKPMWLEVWLCYCKYGRPAEDATSVARTVQNNQSLINYYMYNNGCNGIC 300

QY , 3UI SSGLF IATSIDYDAPLDEYGLLNEPKYGHRLDLHRAIKLSEPALVSSYAAVTSLGNSQEA 360

bb 301 ssglfiatsydydaplddeyglnepkyyghlrdlnhkaiklsepalvssyaavtslgsnqea 360

361 HVYRSKSGACAAFLSNYDSRYSVKVTFFQNRPNLPPWSISILPDCCKTAVYNTAQVNSQSS 420
QY

Db 361 hvyrsksgacaafslsnydsrysvkvtfqnrpynlppwsisilpdcktavntaqvnsqss 420

QY 421 SIKMTPAGGLSWQSYNEETPTADDSDTLTANGLWEQKNVTRDSSDYLWYMTNVNIA SNE 480

Db 421 sikmtpagglswgsyneetptaddsdlttanglwegknvtrdssdylwymtnvniasne 480

QY 481 GFLKNGKDPYLTVMSAGHVLHVFNGLSGTVYGTLDNPKLTYSGNVKLRAGINKISLLS 540

Db 481 qflknqkdpvltymsaqhvlhvfvpqkllsqtvvqtlldnpklltysqnvklraqinkislls 540

0y 541 VSVGI.PNVGVHYDTWNAGVI.GPVTI.SGL.NEGSRNI.AKOKWSYKVGI.KCFST.ST.HSISGSS 600

db

[illegible]

Db 6 ilmlnvlvl---lgswwsfsgtasvshdralivngqrllisgsvhprstpmwpgii 62
Qy 61 QKADGGDLVETVYFWNGHEPSPGKYNFEGRYDLVRFKIMVORAGLVNLRIGPYVCAE 120
Db 63 qkakeggvdiqtvyfnghepggkyyfegrydlvkfkivhqaglyvhrvrgpyacae 122
Qy 121 WNGGFPVWLKYVPGMEFRTNNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAAQIEN 180
Db 123 wnggfpvwlkyvpgisfrtdngpfkaamqfktakivnmkaerlyetqggpilisqien 182
Qy 181 EYGPVWEIGAPGKAYTKWAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCEGFRPN 240
Db 183 eygmewelgapgksyaqaamavgltdgvpwimckqddapdpilnacngfydyfspn 242
Qy 241 KPYKPKMTEVMTGWTYKFGGPIPORPAEDIAFSAVARVQNGSPFNYYMYHGGTFNGRT 300
Db 243 kaykpkwteawtftgpnvprpaedlafsvakfiqggsfinymyhggtgnfrt 302
Qy 301 SSGLFIAFSDYDAPLDEYLLNEPKYGHRLDLKAIKLSEPALVSSYAAVTSLSGNSQEA 360
Db 303 aggpfiatsydydapldeyglirgpkwghldlhraiklcepalvsgdpavtaighqgea 362
Qy 361 HVIYSKSCACAAFLSNYDSRYSVKVTQFONRPNLPWPSISILPDCKTAVYNTAQVNSOSS 420
Db 363 hvfrskagscaafianyqhsfatvtsfanrhnlpwpsisilpdckntvntarigaqsa 422
Qy 421 SIKMTPAGGGLSWOSYNBETPTADDSDTLTANGLEWQKNVTRDSSDYLWYMTNVNVIASNE 480
Db 423 qmktptvsgripwgsfneetsyeds-sftvvglleqinttrdvsdylwystdvwkdsre 481
Qy 481 GFLKNGKDPYLTVMASGHVLHVFNVGKLSGTVYGTLDNPKLTYSGNVKLKLAGINKISLLS 540
Db 482 kflrggkwptlmsaghalhvfngqgiagtaygslekpkltskavnlragvgnkislls 541
Qy 541 VSVGLPNVGHYDTWNAAGLVPTVLISGLNEGRNLAKQWYKVLKCESLSLSLSSSS 600
Db 542 lavgplnlgpfetwngavlgpvsitgidedkrdltwqkwykvgkgealslslsgss 601
Qy 601 SVEWVRSLMAQKOPLTWYKATFNAPGNDPLALDMSMGQGIWINGEGVGRHWPGYIA 660
Db 602 sveewegslvaqrpltwyktfnapagnndplalntmgkgwvngsglrywpgyka 661
Qy 661 QGDCSKSYAGTFNEKKQCTQNCQSPQRWYHVPRSWLKPNSNLLVVFEEGNGNPTGISLV 720
Db 662 sgncgacnyagwfnckclscgeasgrwyhvrswlyptgtullvlfeewggegpgislv 721
Qy 721 RR 722
Db 722 kr 723

RESULT 3
AAY44303
ID AAY44303 standard; protein; 835 AA.

XX -

AC AAY44303;

XX -

DT 29-FEB-2000 (first entry)

DE Tomato beta galactosidase-1.

XX Tomato beta galactosidase-1; TBG; Rutgers tomato plant; pectin;
KW fruit softening; beta galactosidase II protein; biofilm;
KW transgenic plant; protoplast isolation.

XX Lycopersicon esculentum.

OS

XX Key Location/Qualifiers

FH Peptide 1..24

FT /label= Signal_peptide

FT Protein 25..835

XX /label= beta-galactosidase-1

PN W09964564-A1.
XX 16-DEC-1999.
XX 08-JUN-1999; 99WO-US12697.
XX 09-JUN-1998; 98US-0088805.
XX (USDA) US DEPT OF AGRICULTURE.

PI Gross KC, Smith DL;

XX WPI; 2000-097532/08.

DR N-PSDB; AAZ29338.

XX New beta-galactosidases, used to prepare transgenic plants with altered
fruit ripening -

PS Claim 1; Fig 2; 85pp; English.

XX The present sequence is tomato beta galactosidase-1 (TBG-1) encoded by a
CC cDNA derived from breaker, turning and pink fruit pericarp from 'Rutgers'
CC tomato plants. This hydrolyses terminal non-reducing beta-D-galactosyl
CC residues from beta-D-galactosides leading to loss of tissue integrity and
CC fruit softening. This is used for modifying cell wall metabolism and
CC controlling ripening of fruit by altering activity of beta galactosidase
CC II protein. Pectin with reduced galactosyl content is produced for use in
CC biofilms or solutions. Transgenic plants with altered fruit ripening are
CC produced by introducing DNA constructs comprising TBG cDNA. TBG forms a
CC component of an enzyme mixture used to isolate protoplasts.

XX Sequence 835 AA;

Query Match 74.5%; Score 2922.5; DB 21; Length 835;
Best Local Similarity 72.2%; Pred. No. 1.le-239;

Matches 518; Conservative 94; Mismatches 102; Indels 3; Gaps 2;

Qy 6 VLLLVLCLLDPFFSVKASVSYDDRAITINGKRKTLISGSIHVPRSTQMWPDLIQKAD 65

Db 7 mlmlmlcl--wvscgiasvdydhraivngqrllisgslhyprstpmwpdliqkake 64

Qy 66 GGLDVIETVYFWNGHEPSPGKYNFEGRYDLVRFKIMVORAGLVNLRIGPYVCAEFGG 125

Db 65 ggvdvdiqtvyfnghepgkyyfegrydlvkfkivhqaglyvhrvrgpyacaewnfgg 124

Qy 126 FPVWLKYVPGMEFRTNNQPFKVMQGFVQKIVNMKSENLFESQGGPIIMAAQIENYGPV 185

Db 125 fpvwlkyvpgisfrtdngpfkaamqfktakivnmkaerlyetqggpilisqienygm 184

Qy 186 EWEIGAPGKAYTKWAQMAVGLKTGVPWIMCKQEDAPDPVIDTCNGFYCEGFRPNKYPK 245

Db 185 eweigepkvyseawaakmavdlgtgvpwimckqddapdpilntcngfydyftpnkankp 244

Qy 246 KMTEVWTGWYTKFGGPIPORPAEDIAFSAVARVQNGSPFNYYMYHGGTFNGRTSSGLF 305

Db 245 kmwteawtawtftgpgpyrpaedmafavarfiqtggsfinymyhggtngfrtsggpf 304

Qy 306 TATSVDYDAPLDEYLLNEPKYGHRLDLKAIKLSEPALVSSYAAVTSLSGNSQEAHVRS 365

Db 305 latsydydapidefgslrqpkwghlkdhlraiklcepalvsvdptvtslgnqgearvfks 364

Qy 366 KSGACAAFLSNYDSRYSVKVTQFONRPNLPWPSISILPDCKTAVYNTAQVNSOSSIKMT 425

Db 365 esgacaaflanyqhsfakvafgmhynlpwpsisilpdckntvntarvgaqsgmqmt 424

Qy 426 PAGGLSQSQSYNEETPTADDSDTLTANGLEWQKNVTRDSSDYLWYMTNVNVIASNEGFLKN 485

Db 425 pvsrgfswesfnedaashed-dftvvgllqeqinrtvdsdylwmtdeidptegfins 483

Qy 486 GKDPYLTVMASGHVLHVFNVGKLSGTVYGTLDNPKLTYSGNVKLKLAGINKISLSVSVGL 545

Db 484 gnwpwltvtsaghalhvfngqgiagtaygslenpkltsfnginlragvgnkisllsiavgl 543

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162143.

Query Match 71.68; Score 2808.5; DB 21; Length 724;
Best Local Similarity 70.2%; Pred. No. 4.4e-230;
Matches 503; Conservative 92; Mismatches 119; Indels 3; Gaps 3;
Qy 7 LLLVLCILDFSSVKASVSDRAIINGKRKILISGSIHYPRSTPOMWDLIOKAKDG 66
Db 11 iflailccclslscivkasvsydrkavlingqrlllsgsihyprstpmwpgliqakeg 70
Qy 67 GLDVIETYVFWNGHEPSPGKYNFEGRYDLVRFIKMWQRAGLYVNLRIIGYVYCAEWNFGGF 126

Db 71 gldvetyvfnghpepggyfdrydlvklfihvdaglyvnlrigrpyycaewngfgf 130
Qy 127 pwlklyvpgmefrnnopfkvamogfvyokivnmkksenlfesogqpiimaioieneygpve 186
Db 131 pwlklyvpgmafrtdnepfkamkfkfekivmmmaeklfqgqgpiilaqieneygpve 190
Qy 187 weigapgaytkwaamaavglktgvpwimckoeapdpvldtcngfycegrprnkpkykp 246
Db 191 weigapgaytkwaamaavglktgvpwimckoeapdpvldtcngfycegrprnkpkykp 250
Qy 247 mtevwgwytkfkgpripdorpaeadiarsvarfvqngsffnyvmyhgtngftrssglfi 306
Db 251 mwtenwtgwytdfggavpyrpediayvarfiqggslnvymyhggtntfdr- agefm 309
Qy 307 atsydydapldexgllnpepkghlrdlhkaiklsepalsvssaaavtslgsnoeahvyrsk 366
Db 310 assydydapldexgllnpepkghlrdlhkaiklsepalsvssaaavtslgsnoeahvyrsk 369
Qy 367 sgacaafslsnydsrysvkvtfqonrpylnppwsiilpdkctavxyntaqvnsqssikmtp 426
Db 370 s-scaafslsnydsrysvkvtfqonrpylnppwsiilpdkctavxyntaqvnsqssikmtp 428
Qy 427 agglswsqsyneerptaddstflranglweoknvtrossdylwtvntvniassnegflkng 486
Db 429 tgtkfsgsfneatpneagtfnarglveqslmtwdksdyfwyitdtigsetflktg 488
Qy 487 kdpvltvmsaghlvfvngklsgtvygtldnplklytsnkvklraginkisllsvsglpl 546
Db 489 dsplltvmsaghlvfvngklsgtvygtldnplklytsnkvklraginkisllsvsglpl 548
Qy 547 nvghydtvwnagvlpgvptlsglnegrnlakokvskvklgkgeslshslsgssvewvr 606
Db 549 nvghydtvwnagvlpgvptlsglnegrnlakokvskvklgkgeslshslsgssvewvr 608
Qy 607 gslmaqoplwytkatfnagpndplaldmasmkkgiwtwngvgrhwpqyiaogdcsk 666
Db 609 gsfvakkpqltwytkatfnagpndplaldmasmkkgiwtwngvgrhwpqyiaogdcsk 668
Qy 667 cysagtfnkctocngcoporvhyvprswlksplkpsgllvfpewgngptglsllvrrs 723
Db 669 cnyagtfdaakclscnsgearvhyvprswlksplkpsgllvfpewgngptglsllvrrs 724

RESULT 7
AAG49310
ID AAG49310 standard; Protein; 666 AA.
XX AC AAG49310;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 62369.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126284.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.


```

|| || || | :|:||||: ||| :|||: ||||| ||||| | ||:|||| ||:|
Db 599 wpaqagsgcncryagtfdakcklscnqceasgrwyhvrswlk-sgnlivvfeelggdp 657
QY 715 TGISLVRRS 723
    |||||:|
Db 658 ngislvrkt 666

RESULT 8
AAG49423
ID AAG49423 standard; Protein; 853 AA.
XX
AC AAG49423;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62523.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
```

```
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 66.5%; Score 2608; DB 21; Length 853;
Best Local Similarity 63.1%; Pred. No. 6.9e-213;
Matches 457; Conservative 119; Mismatches 136; Indels 12; Gaps 5;
```

```
OY 8 LLLVLICL-----LDFSSVAKASVYDDRAIIINGKRKTLISGSIHYPRSTPQMPDLIIQ 61
Db 10 lllwclglflilgvvf---vqcgvydrkallingqrllfsgsihyprstpdmedliq 66
```

```
OY 62 KAKDGLDVIETYYVFWNGHEPSPGKYNFGRYDLVRFIMQVORAGLYVNLRTGPPYCAEW 121
Db kakkdgldvietyvnllhhepspgkydfegrnldvrfvktihkaglyahrltgyvcaew 126
OY 122 NFGGPPVWLKYYPGMEFFRTNNOPEKVMQGFVOKTVNMKMSLENLPESSOGPIIQAQIENE 181
Db nfggppvwlkyyvgisfrtdnepfkramkgfterivelmkslenlfesggppqllsqiene 186
OY 182 YGPVEWEIGAPGKAYTKWAAQMAVGLKTGPWIMCKQEDADPVPIDTCNGFYCEGFRPNK 241
Db ygpveweigapgkaytkwaaqmaivglktgvpwimckqedadpvpidtcngfyccegrpnk 246
OY 242 PYXPKMTEVWTGWYTKFGGPIQORPAEDIAFSAFVONNGSFYNYMYHGGTNGFRGTS 301
Db pyxpkmtevwgtgwtkfggpiqorpaediafsvafvonnsgsfynymyhggtngfrgts 306
OY 302 SGLFIATSYDYDAPLDEYGLLNEPKYGHRLDLHKAIKLSEPALVSSYAAVTSILGNSQEAH 361
Db sglfiatsydydapldeygllnepkylghrldlhkaiklsepalvssyaaavtsilgnsqeah 366
OY 362 VYRSKSGCAAFSLNVDTRYSVKVTQFONRPNYLPWMSILPDCKTAVYNTAQNVSQSS 421
Db vyrsksgcacaafslnvdtrysvkvtqfonrpnynlpwmsilpdcktavyntaqnvsqss 426
OY 422 IKMTPAG-GGLSQSYNEETPTADSDTLTANGLEQKNVTRDSSDYLMYMTNVNIASNE 480
Db ikmtpag-gglsqsyneetptadssdtltangleqknvtrdssdylmymtnvniasne 486
OY 481 GFLKNGKDPYLTVMAGHVLHVFNAGKLSGTYYGTLDPKLTYSGNVKLRAGINKISLLS 540
Db gflkngkdpyltvmaghvlhvfnagklsgtyygtldpkltysgnvklraginkislls 546
OY 541 VSVGLPNVGHVYDVTNAGVLGPTLISGLNRSNLAKQSKYKVLKGBSLSLHSLSGSS 600
Db vsvglpnvghvydvtnagvlgptlislglnsnlakqskykvlkgsllshslsgss 606
OY 601 SVEVYRGSLSMAQK-QPLTWYKATFNAPGNDPLALDMSMGKGQIWIINGEGVRRHWPGYI 659
Db sveyvyrsgslmaqk-qpltwykatfnapgndplaldmsmgkgqiwiingegvrrhwpgyi 665
OY 660 AOGDCSKSYAGTFNEKKCOTNCGOPSORWYHVPKSLKPSGNLAVPEEMGNGNPTGISL 719
Db aogdcsksyagtfnekkcotncgopsorwyhvpkslkspsgnlavpeemgngnptgisl 725
OY 720 VRRS 723
Db vrrs 729
OY 726 VRRS 729
Db vrrs 729

RESULT 9
AAG49422
ID AAG49422 standard; Protein; 856 AA.
XX
AC AAG49422;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 62522.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
```


PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148664.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match		65.5%; Score 2572; DB 21; Length 847;
Best Local Similarity		62.0%; Pred. No. 8e-210;
Matches 453; Conservative 118; Mismatches 128; Indels 32; Gaps 6;		
QY	8 LLLVCL-----LDFFSVKASVSDRAIIINGKRKILISGSIHYPSPQWMPDLIQ	61
Db	10 lllwclglilvgf---vqcgvydrkallingrrilfsgsihprstpdmwdliq	66
QY	62 KAKDGLDVIETYPVWNGHEPSPGKYNEGRYDLVRFETKMQVQAGLYVNLRIQPVVCAEW	121
Db	67 kdkdgiadietvfnlhhepspkdyfegndlvrfvktlkhgilyanhrigpyvcae	126
QY	122 NFGGPPVWLKYPGMEFRTNQPKVAMQGVQKIVNMKMSLENFESOGGPIIIMAAIQENE	181
Db	127 nfggfpvwlkypgisfrtdnepkramkqfteriveimkseenfesggpillsqiene	186
QY	182 YGPWEWEIGAPGKAYTKWAAQVGLKTGVPMWCKQBDADPPVIDTCNGYCEGFRPNK	241
Db	187 ygrqgqligaeghnymtwaakmatetgvpwvmckedapdpvintcngfydsfapnk	246
QY	242 PYKPMWTEVMTGWYTKFGGPIPORPAEDIAFSVARFVQNNGSFENYMYHGGTFNFGRTS	301
Db	247 pykplwteawsgwftfegpphhprvqdlafgvarfikggsfvnyymyhggtafgrta	306
QY	302 SGLFATSYDADPLDEYGLLNEPKYGHRLDLKAIKLSEPALVSSYAATVSLGSNQF--	359
Db	307 ggpfttsydydapidvylirqpkyghikelhraikncekalysadpvtvtsignkqgvw	366
QY	360 -----AHVYRSKSGACAAFLSNDTSRYSVKVTQNRPNYLPWWSILPLDCKTAVYNTA	413
Db	367 iyyerfahvysaesgdcsafianydtesaarvlnnvhynlpwvssilpdcnnavfnta	426
QY	414 QVNSQSSIKWTPAGGGLSWQSYNEETPTADSDTLTANGLWEOKNVTRODSSDYLWYMTN	473
Db	427 kvsn-----fqwesyledssldsdstftchleginvtrcdtsdylwymts	473
QY	474 VNIASENEGFLNKDYPYLVNAGHVLHVYVNGKLSGFVYGTLDNPKLITYSGNVKLRAGI	533
Db	474 vdigdsesflhgelpitliigstghavhifvngqlsgsafqrnrrftygkkinlhsqt	533
QY	534 NKISLLSVSGLPNVGHYDTPWAGVLPVTLGSLNEGSRNLAKQKSYKVLKGESLSL	593
Db	534 nriallsavaglpnvghfeseswtgilpvalhglsgqkmdslwqkwtvgvlgkgeamnl	593
QY	594 HSLGSSSVWVRGSLMAOK-QPLTWYKATFNAPGNDPLALDMSAKGGOIWIINGEGVG	652
Db	594 afptntpsigmndasltvqkppqtlwhtkyfdaepgneplaldmegmkqgiwvngesig	653
QY	653 RHPGPIYIAGDCSKCSYAGTNEKKCQTPNCGQPSQRYHVPRLKPSGNLLVVFEEWGG	712
Db	654 rywtaf-atgdchscsytygtykpnkcqtcgqgptqrwyhvprawlkpsqnlvifeelgg	712
QY	713 NPTGSLVRRS 723	
Db	713 npstsvlvkrs 723	
RESULT 11		
AAG31161		
ID	AAG31161 standard; Protein: 850 AA.	
XX		
AC	AAG31161;	
XX		
DT	17-OCT-2000 (first entry)	
XX		
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 37376.	
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway;	
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	
KW	termination sequence.	
XX		
OS	Arabidopsis thaliana.	

XX	EP1033405-A2.
PN	
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126284.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.
PR	23-APR-1999; 99US-0130510.
PR	28-APR-1999; 99US-0130891.
PR	30-APR-1999; 99US-0131449.
PR	30-APR-1999; 99US-0132048.
PR	04-MAY-1999; 99US-0132407.
PR	05-MAY-1999; 99US-0132484.
PR	06-MAY-1999; 99US-0132485.
PR	06-MAY-1999; 99US-0132486.
PR	07-MAY-1999; 99US-0132487.
PR	11-MAY-1999; 99US-0132863.
PR	14-MAY-1999; 99US-0134256.
PR	14-MAY-1999; 99US-0134218.
PR	14-MAY-1999; 99US-0134219.
PR	14-MAY-1999; 99US-0134370.
PR	18-MAY-1999; 99US-0134768.
PR	19-MAY-1999; 99US-0134941.
PR	20-MAY-1999; 99US-0135124.
PR	24-MAY-1999; 99US-0135353.
PR	25-MAY-1999; 99US-0136021.
PR	27-MAY-1999; 99US-0136392.
PR	28-MAY-1999; 99US-0136782.
PR	01-JUN-1999; 99US-0137222.
PR	03-JUN-1999; 99US-0137528.
PR	04-JUN-1999; 99US-0137502.
PR	07-JUN-1999; 99US-0137724.
PR	08-JUN-1999; 99US-0138094.
PR	10-JUN-1999; 99US-0138540.
PR	10-JUN-1999; 99US-0138847.
PR	16-JUN-1999; 99US-0139119.
PR	16-JUN-1999; 99US-0139452.
PR	16-JUN-1999; 99US-0139453.
PR	17-JUN-1999; 99US-0139492.
PR	18-JUN-1999; 99US-0139454.
PR	18-JUN-1999; 99US-0139455.
PR	18-JUN-1999; 99US-0139456.
PR	18-JUN-1999; 99US-0139457.
PR	18-JUN-1999; 99US-0139458.
PR	18-JUN-1999; 99US-0139459.
PR	18-JUN-1999; 99US-0139460.
PR	18-JUN-1999; 99US-0139461.
PR	18-JUN-1999; 99US-0139462.
PR	18-JUN-1999; 99US-0139463.
PR	18-JUN-1999; 99US-0139750.
PR	18-JUN-1999; 99US-0139763.
PR	21-JUN-1999; 99US-0139817.
PR	22-JUN-1999; 99US-0139899.
PR	23-JUN-1999; 99US-0140353.
PR	23-JUN-1999; 99US-0140354.
PR	24-JUN-1999; 99US-0140695.
PR	28-JUN-1999; 99US-0140823.
PR	29-JUN-1999; 99US-0140991.
PR	30-JUN-1999; 99US-0141287.

```
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.5%; Score 2572; DB 21; Length 850;
Best Local Similarity 62.0%; Pred. No. 8e-210;
Matches 453; Conservative 118; Mismatches 138; Indels 32; Gaps 6;

Qy 8 LLLVICL-----LDFFSSVKASVSYDDRAIIINGKKILISGSIHPRSTPQMWDLIQ 61
Db 10 lilwclgfililvgvf---vcgvtydrkallingqrilfsgsihyprstpdmdwliq 66
Qy 62 KAKDGLDVIETYYFVWNGHEPSPCKYNFEGRYDLVRFTKMWORAGLVNLRIGPYVCAEW 121
Db 67 kagdggladvietyvfnlhhepspgkydfregndlvrvktlnkaglyahnlrigpyvcaew 126
Qy 122 NFGGPPVWLKYVPGMEFRTNNQPKVAMQGFQKIVNMKSENLFESOGGPIIMAOIENE 181
Db 127 nfggfpwlkyvpgisfrtdnepfkramkgfterivelmksefnifesgggpiilsiqene 186
Qy 182 YGPVENEIGAPGKAYTKWAAQMAVGLTGVPMCKEDADDPDVIDTCNGFYC6GFRPNK 241
Db 187 ygrggqllgaeghnymtwaaakmatetgvpwmckeddapdvintcngfydcfsfapnk 246
Qy 242 PYKPKMTEVWTGTYTKFGGPIPORPAEDIAFSVARFVQNNNGSNFYMYHGGTNFGRTS 301
Db 247 pykplwteaawsgwftgpgmhhrrpvqdlafgvarfiqkggsfvnyymyhggnfgrta 306
Qy 302 SGLFIATSYDYDAPLDEYGLNPEKYGHLRDLHKAIKLSEPALVSSYAAVTSLSGNOE-- 359
Db 307 ggpfttsydydapideyglirqpkyghlkelhraiikmcekalysadvpvttsignkqvw 366
Qy 360 -----AHVYRSKSGACAAFLSNYDSRYSVKVTQFONRPNYLPNPPWSISILPDCKTAVYNTA 413
Db 367 iyyerfahvysaesgdcsaflnydtesaarvlfnvhnlpwpsisilpdcnrvnta 426
```


PR	27-JUL-1999;	99US-0145913
PR	27-JUL-1999;	99US-0145918
PR	27-JUL-1999;	99US-0145919
PR	28-JUL-1999;	99US-0145951
PR	28-AUG-1999;	99US-0146386
PR	02-AUG-1999;	99US-0146388
PR	02-AUG-1999;	99US-0146389
PR	03-AUG-1999;	99US-0147038
PR	04-AUG-1999;	99US-0147202
PR	05-AUG-1999;	99US-0147302
PR	05-AUG-1999;	99US-0147192
PR	06-AUG-1999;	99US-0147260
PR	06-AUG-1999;	99US-0147303
PR	09-AUG-1999;	99US-0147416
PR	09-AUG-1999;	99US-0147493
PR	09-AUG-1999;	99US-0147935
PR	10-AUG-1999;	99US-0148171
PR	11-AUG-1999;	99US-0148319
PR	12-AUG-1999;	99US-0148341
PR	13-AUG-1999;	99US-0148565
PR	13-AUG-1999;	99US-0148684
PR	16-AUG-1999;	99US-0149368
PR	17-AUG-1999;	99US-0149175
PR	18-AUG-1999;	99US-0149426
PR	20-AUG-1999;	99US-0149722
PR	20-AUG-1999;	99US-0149723
PR	20-AUG-1999;	99US-0149929
PR	23-AUG-1999;	99US-0149902
PR	23-AUG-1999;	99US-0149930
PR	23-AUG-1999;	99US-0150566
PR	26-AUG-1999;	99US-0150884
PR	27-AUG-1999;	99US-0151065
PR	27-AUG-1999;	99US-0151066
PR	27-AUG-1999;	99US-0151080
PR	30-AUG-1999;	99US-0151303
PR	31-AUG-1999;	99US-0151438
PR	01-SEP-1999;	99US-0151930
PR	10-SEP-1999;	99US-0152363
PR	10-SEP-1999;	99US-0153070
PR	13-SEP-1999;	99US-0153758
PR	15-SEP-1999;	99US-0154018
PR	16-SEP-1999;	99US-0154039
PR	20-SEP-1999;	99US-0154779
PR	22-SEP-1999;	99US-0155139
PR	23-SEP-1999;	99US-0155486
PR	24-SEP-1999;	99US-0155659
PR	28-SEP-1999;	99US-0156458
PR	29-SEP-1999;	99US-0156596
PR	04-OCT-1999;	99US-0157117
PR	05-OCT-1999;	99US-0157753
PR	06-OCT-1999;	99US-0157865
PR	07-OCT-1999;	99US-0158029
PR	08-OCT-1999;	99US-0158232
PR	12-OCT-1999;	99US-0158369
PR	13-OCT-1999;	99US-0159293
PR	13-OCT-1999;	99US-0159294
PR	14-OCT-1999;	99US-0159295
PR	14-OCT-1999;	99US-0159329
PR	14-OCT-1999;	99US-0159330
PR	14-OCT-1999;	99US-0159331
PR	14-OCT-1999;	99US-0159637
PR	14-OCT-1999;	99US-0159638
PR	18-OCT-1999;	99US-0159584
PR	21-OCT-1999;	99US-0160741
PR	21-OCT-1999;	99US-0160767
PR	21-OCT-1999;	99US-0160768
PR	21-OCT-1999;	99US-0160770
PR	21-OCT-1999;	99US-0160814
PR	21-OCT-1999;	99US-0160815
PR	22-OCT-1999;	99US-0160980
PR	22-OCT-1999;	99US-0160981
PR	22-OCT-1999;	99US-0160989
PR	25-OCT-1999;	99US-0161404

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134250.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142300.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0132407.
PR 21-JUL-1999; 99US-0132484.
PR 22-JUL-1999; 99US-0132485.
PR 22-JUL-1999; 99US-0132486.
PR 22-JUL-1999; 99US-0132487.
PR 22-JUL-1999; 99US-0132863.
PR 23-JUL-1999; 99US-0134250.
PR 23-JUL-1999; 99US-0134218.
PR 23-JUL-1999; 99US-0134219.
PR 26-JUL-1999; 99US-0134276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

```
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      63.3%; Score 2482.5; DB 21; Length 794;
Best Local Similarity 64.1%; Pred. No. 3e-202;
Matches 430; Conservative 110; Mismatches 128; Indels 3; Gaps 3;

Qy 55 MWPLIQAKDGLDVTETVFNWNGHSPKYNPEGRYDLVRITKMWQVQAGLVNLRIG 114
Db 1 mmedliqakdgdietyvfnwlnhepsdkdyfegndlvrvfktihkaglyahlrig 60

Qy 115 PYVCAENFGGPPVWLVKVPCEMTNNOPEKVMQGVQKIVNMKSENLFESGGPII 174
Db 61 pyvcaenfggppvwlkypvpsirtdnepkframkqfterivelmksenlfesgggpii 120

Qy 175 MAQIENYGPVEWETGAPKAYTKWAQMAVGLKTGVPWIMCKQEDADPDVIDTCNGFYC 234
Db 121 lsqienygpvewetgagkaytkwaqmayglktgvpwimckqedadpdpvidtcngfyc 180

Qy 235 EGFRPNKYPKPMKTEVWNTGYTKFGGIPORPAEDIAFSAVFQVQNGSNFFNYMYHGG 294
Db 181 dsfapnkypkplwteawsgwftfeggmhbrpvdlaafvarfiqkgstfvnyymyhg 240

Qy 295 TNFGRTSSGLFIANSYDAPLDVGLLNEPKYGHRLDKAIIKLSPEALVSSAAVNTSL 354
Db 241 tnfgrtsgglfiansydydapldvlgllnepkylghrldkaiiklspealvssaaavntsl 300

Qy 355 GSNQEAHYRSKSGACAFSLNYSRYSKVTFFONRPNLPPWISILPDCKTAVYNTAQ 414
Db 301 gnkqahvysaesdcasflanydtesaarvlfnnvhylnlppwsisilpdcrnafvntak 360

Qy 415 VNSQSSSTKWTAPAG-GLLSWQSYNEETPTADDSDTLFANGLWEOKNVTRDSSDYLWYMN 473
Db 361 vgvqtsqmempitdtknfqwesyledissldsdstftthgleginvtrdtsdylwymts 420

Qy 474 VNIASEGFLKNGKDPYLTVMAGHLVHFVNGKLSGVYCTLDNPKLTVSGNVKLRAGI 533
Db 421 vdiqdsesflhgaelptliqstghavhifvngqlsgsafqrnrrftygqklnlhsgt 480

Qy 534 NKISLLSVGLPNVGHYDFTWAGVLGPVTLISLINEGRNLAKQKWSYKVKLGESLSL 593
Db 481 nrllallsvaglpnvghfswmngilgpvalhlgsgkmdlswqkwtqyqglkgeamnl 540

Qy 594 HSLSGSSVVEVRGSLMAQK-OPITWTWKATENAPGNDPLDALDMASMGKGQIWIINGEGVG 652
Db 541 afptntpsigwmdasiltvqkbpqitwhktyfdapegneplaldmegmgkgiwnvngesig 600

Qy 653 RHPFGYIAQGDSCSYAGTNEKKQTCNCGOPQORWVHVRSLKPSGNLLVFEWGG 712
Db 601 rwtatf-atgdcshcsytgtkpnkcgqcgqptqrwyhvprawlkpqsnllvifeelgg 659

Qy 713 NPTGISLVRSS 723
Db 660 npstsvlvkrs 670
```

RESULT 15

```
AAG39972
ID AAG39972 standard; Protein; 788 AA.
XX
AC AAG39972;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49532.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
```

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145142.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146385.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161902.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 62.3%; Score 2446.5; DB 21; Length 788;

Matches 426; Conservative 109; Mismatches 120; Indels 23; Gaps 4;

Qy 55 MWPLIQAKDGLDVIETVYFWNGHSPSGKYNFEGRYDLVRFTKMVORAGLYVNLRIg 114
Db 1 mwedliqakdggidvetyvfwnlhepspgkydfegndlvrfvktlkhaglyahlrig 60
Qy 115 PYCAEWNFGGFPYWLKYVPGMEPTNNQPKVAMOGVQKIVNMKSENLFESOGGPii 174
Db 61 pyvcaewnfggfpwykvyvpgisfrtdnepfkramkgfiterivelmksenlfesgggpil 120
Qy 175 MAQIENEXGPPVEWEIGAPGKAYTKWAQMAVGLKTGVPWIMCKQEDADPPVIDTCNGFYC 234
Db 121 lsqieneygrgqqlgaeghnymtwaakmaiatetgvpwvnmckeddapdpvintcngfyc 180
Qy 235 EGFRPNKPKPKMTEVWTGWYTKFGGPiPORPAEDIAFVSARFVQNNNGSFFNYMYHGG 294

[illegible]

Search completed: June 24, 2002, 20:38:20
Job time: 155 sec

THIS PAGE BLANK (USPTO)